



Synthèse sonore d’ambiances urbaines pour les applications vidéoludiques

Guillaume Tiger

► To cite this version:

Guillaume Tiger. Synthèse sonore d’ambiances urbaines pour les applications vidéoludiques. Acoustique [physics.class-ph]. Conservatoire national des arts et metiers - CNAM, 2014. Français. NNT : 2014CNAM0968 . tel-01149306

HAL Id: tel-01149306

<https://theses.hal.science/tel-01149306>

Submitted on 6 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT

Présentée par : **Guillaume TIGER**

Soutenue le : **2 décembre 2014**

Pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/spécialité : **Informatique**

Synthèse sonore d'ambiances urbaines pour les applications vidéoludiques

Thèse dirigée par :

M. Stéphane NATKIN

Professeur – Cnam

Codirection :

Mme Cécile LE PRADO

Maître de conférences associée – Cnam

Rapporteurs :

Mme Rozenn NICOL

Ingénieur de recherche – Orange Labs, HDR

M. Geber L. RAMALHO

Professeur – Université fédérale de Recife – Brésil

Examineurs :

M. Alexandre GARCIA

Professeur – Cnam

M. Dominique ARCHAMBAULT

Professeur – Université Paris 8

M. Antoine GONOT

Maître de conférences – Aix-Marseilles Université – Dép.SATIS

M. Diemo SCHWARZ

Chercheur – IRCAM – Paris

When I hear what we call music, it seems to me that someone is talking. And talking about his feelings, or about his ideas of relationships. But when I hear traffic, the sound of traffic – here on Sixth Avenue, for instance – I don't have the feeling that anyone is talking. I have the feeling that sound is acting. And I love the activity of sound [...] I don't need sound to talk to me.

John Cage, entretien avec Miroslav Sebestik, *Listen*¹

Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts... A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the non space of the mind, clusters and constellations of data. Like city lights, receding...

William Gibson, *Neuromancer*, Ace Books, 1984, p.69

1. Source : «Listen», documentaire de Miroslav Sebestik, ARTE France Développement, 2003.

Résumé

Les ressources mémoire et processeur allouées au rendu audio des applications vidéoludiques sont une contrainte importante dans le processus de réalisation d’ambiances sonores complexes. Des compromis sont nécessaires dans le choix des matériaux sonores et de leurs traitements pour parvenir à un résultat temps-réel crédible et immersif. Des méthodes alternatives dites *procédurales* tendent à être de plus en plus utilisées, elles consistent à modeler le matériau sonore requis en fonction d’informations récupérées dans l’environnement virtuel. Les méthodes *procédurales* semblent appropriées à la sonorisation d’environnements virtuels complexes et évolutifs tels que des environnements urbains.

Dans ce travail de thèse, nous nous intéressons tout particulièrement à l’élaboration d’ambiances sonores urbaines interactives. Nous définissons la composition de ces ambiances au travers des analyses proposées par la théorie du Paysage Sonore et d’un état de l’art sur les applications vidéoludiques mettant en scène de vastes environnements urbains. Nous en retirons une segmentation du paysage urbain virtuel en sous-ensembles de plans sonores perceptifs, incluant un arrière-plan. L’arrière-plan se distingue par son ubiquité, sa stabilité spectrale dans le temps et sa capacité à renseigner sur l’état de l’environnement urbain pris dans son ensemble.

Nous considérons ainsi l’arrière-plan sonore urbain virtuel comme une *texture* et proposons un outil de resynthèse granulaire dédié à sa sonorisation interactive. Cette thèse a pour objectif de définir, spécifier et prototyper un outil de synthèse de l’arrière-plan urbain utilisable dans des applications fortement interactives telles que le jeu vidéo. Cet outil est principalement articulé autour de lecteurs granulaires dont les sorties sont filtrées et mixées dynamiquement relativement à des paramètres transmis par l’environnement virtuel. Nous détaillons la réalisation du synthétiseur granulaire dans l’environnement de pro-

grammation graphique Pure Data ainsi que son fonctionnement. Sur la base de l'état de l'art, nous accompagnons cette proposition d'une méthodologie d'enregistrement des ambiances urbaines afin d'alimenter la resynthèse.

Le développement du synthétiseur d'arrière-plans sonores urbains est réalisé dans le contexte du projet de simulation Terra Dynamica proposant une représentation tridimensionnelle et navigable de Paris. Terra Dynamica met en avant deux des principales problématiques de la sonorisation de scènes urbaines : la densité des acteurs présents dans la scène et la variabilité de position du point d'écoute. La segmentation des scènes virtuelles en différents plans apporte une solution à la sonorisation simultanée de plusieurs centaines d'acteurs, en nous basant sur une analyse temps-réel du contenu de la scène et de la position de la caméra virtuelle il nous est possible de faire varier les principaux paramètres de la resynthèse granulaire.

Nous complétons la validation du prototype par un test perceptif d'écoutes comparatives. Les résultats de ce test tendent à montrer que notre méthode, appliquée aux arrière-plans sonores urbains, n'est pas moins bonne que la méthode de synthèse concaténative à laquelle nous la comparons.

Les intégrations et tests menés sur notre prototype tentent à démontrer son efficacité à produire des arrière-plans sonores variés et dynamiques. Dans l'ensemble de ce travail, l'arrière-plan sonore apparaît comme un élément à part entière de l'environnement virtuel de part ses caractéristiques morphologiques et les informations qu'il communique, tandis qu'aucun outil procédural dédié à sa sonorisation ne semble préexister. Nos dernières expérimentations avec le synthétiseur, y compris dans le moteur Unity3D, laissent à penser que le prototype proposé pourrait participer des prémisses d'un tel outil.

Mots clefs : synthèse granulaire, audio procédural, Pure Data, paysage sonore urbain, bruit de fond, jeu vidéo, design sonore, field recording

Remerciements

Je remercie sincèrement toutes celles et ceux dont les encouragements, les idées et la présence ont contribué à la réalisation de cette thèse...

Stéphane Natkin, Cécile le Prado, Shih-Han Chan, Guillaume Levieux, Lubna Odeh, Stéphanie Mader, Jérôme Dupire, Delphine Soriano, et toute l'équipe ILJ, sans oublier Alexandre Topol.

Elizabeth Maler et Raphaël Robert-Bouchard. Romain Arbez, Romain Bathelemy et Maxime Gillet. Damien Henry et AudioGaming. Pascal Peyronnet et l'équipe de Terra Dynamica.

Rozenn Nicol et Geber Ramalho.

David Christoffel.

Coralie.

Trames sonores : John Cage - *In a landscape*, György Ligeti - *Lontano*, Thomas Köner - *Novaya Zemlya*, Tristan Murail - *Légendes urbaines*, Brian Eno - *Ambient 4 : On Land*, Andrew Thomas - *Between buildings and trees*, Gas - *Nah Und Fern*, John Luther Adams - *At a still point*, Henri Dutilleul - *Tout un monde lointain*, Robert Rich - *Illumination*

Table des matières

1	Introduction générale	15
1.1	Contexte de la thèse : le projet Terra Dynamica	18
1.2	Méthodologie	19
1.3	Plan de la thèse	20
2	État de l’art : Espaces sonores urbains	22
2.1	Concepts fondamentaux	24
2.1.1	Paysages, écologie, environnements, scènes sonores	25
2.1.1.1	Paysage et écologie sonores : des outils d’analyse	28
2.1.1.2	Environnements sonores	32
2.1.1.3	Le paysage sonore comme méthodolo- gie de création	33
2.1.2	Timbres sonores	36
2.1.2.1	Espaces de timbres	36
2.1.2.2	Timbre et sons environnementaux . . .	37
2.2	La ville sonore	39
2.2.1	Analyse acoustique	40
2.2.1.1	Champs sonores	41
2.2.1.2	Spectre de l’arrière-plan urbain	43
2.2.2	Analyse perceptive	45
2.2.2.1	Complexité perceptive de la ville sonore	45
2.2.2.2	Objets sonores urbains	47
2.2.2.3	Présence de l’arrière-plan sonore . . .	50
2.3	Virtualisation et sonorisation de la ville	53
2.3.1	Villes sonores au cinéma	53
2.3.1.1	<i>City symphonies</i>	55
2.3.1.2	Villes et ambiances	57
2.3.2	Villes sonores interactives	59

2.3.2.1	Frontières diégétiques des environnements virtuels	60
2.3.2.2	Gestion des arrière-plans sonores dans les jeux vidéo	61
2.3.2.3	Conclusion	68
2.4	Outils de sonorisation des villes virtuelles	72
2.4.1	Synthèse en temps réel d'ambiances urbaines	72
2.4.1.1	Synthèses environnementales	73
2.4.1.2	FM et tables d'ondes	75
2.4.1.3	Synthèse granulaire	77
2.4.1.4	Bilan	81
2.4.2	Outils dédiés au jeu vidéo	82
2.4.2.1	Fmod et Wwise	84
2.4.2.2	Audio procédural et jeux vidéo	86
2.5	L'arrière-plan urbain comme texture sonore	89
2.5.1	Définition d'une texture sonore	89
2.5.2	Synthèse de textures	91
2.5.3	Arrière-plans sonores : textures complexes	94
2.6	Conclusion	96
3	Problématiques de recherche	99
3.1	Introduction	100
3.2	Trois problématiques	101
3.2.1	Quelles articulations pour les plans sonores ?	101
3.2.2	Quelles méthodes de synthèse sonore pour caractériser une ville ?	102
3.2.3	Quelles méthodes de rendu sonore sémantique ?	103
3.3	Points de vue sur l'arrière-plan sonore	105
3.3.1	Le point de vue acoustique	105
3.3.1.1	Paris, de la station Quatre Septembre à Notre Dame de Lorette	108
3.3.1.2	Belgique, Liège	109
3.3.1.3	Paris, Bibliothèque Nationale	110
3.3.1.4	États-Unis, San Francisco	111
3.3.1.5	Conclusion	112
3.3.2	Le point de vue perceptif	112
3.3.3	Le point de vue de la composition et de la mise en scène	114
3.3.4	Dynamique de l'arrière-plan	117

3.3.4.1	Positionnement de l'écoutant et changement de plan pour un objet sonore	117
3.3.4.2	Dynamique de la composition	118
3.3.4.3	Arrière-plan et objets singuliers	119
3.4	Conclusion	121
3.4.1	Caractéristiques de l'arrière-plan sonore urbain	121
3.4.2	Enregistrement de l'arrière-plan sonore urbain	122
4	Réalisation du synthétiseur	124
4.1	Introduction	125
4.2	Synthétiseur d'ambiances urbaines : objectifs et contraintes	126
4.2.1	Outil de design sonore	126
4.2.2	Fonctionnement en temps réel	127
4.2.3	Compatibilité avec Terra Dynamica	128
4.3	Pure Data	130
4.3.1	Historique et versions	130
4.3.2	Pure Data et interactivité	131
4.3.3	Pure Data comme environnement de développement	132
4.4	Architecture	133
4.4.0.1	Exposé théorique de notre approche granulaire	135
4.4.1	Oscillateurs granulaires	137
4.4.1.1	<i>Grainer</i> - Lecteurs de grains	138
4.4.1.2	<i>Blender</i> - Navigation et filtrage	140
4.4.2	Mixeur	142
4.4.2.1	Mixage des oscillateurs	142
4.4.2.2	Rendu au format Ambisonique	143
4.5	Interfaces	145
4.5.1	Paramètres haut-niveau	145
4.5.1.1	Atténuation	145
4.5.1.2	Granularité	146
4.5.1.3	Position dans la table d'ondes	146
4.5.2	Paramètres bas-niveau	146
4.5.2.1	<i>Blend Factor</i>	147
4.5.2.2	Ratio des fondus enchaînés entre les grains	147
4.5.2.3	Calibration des filtres	147
4.5.3	Utilisation d'un contrôleur externe	148
4.6	Construction de la banque sonore de référence	150

4.6.1	Matériel d'enregistrement	150
4.6.2	Matériel d'édition	151
4.7	Conclusion	153
5	Validation et expérimentations	155
5.1	Intégration au projet Terra Dynamica	157
5.1.1	Conception sonore pour le projet	157
5.1.2	Réalisation	160
5.1.2.1	Utilisation de Fmod	160
5.1.2.2	Intégration du synthétiseur	165
5.1.3	Bilan et conclusion	169
5.2	Écoutes comparatives	170
5.2.1	Expérimentation	171
5.2.1.1	Choix des échantillons pour l'expé- rimentation	172
5.2.1.2	Protocole	173
5.2.2	Présentation et analyse des résultats	174
5.2.3	Discussion et conclusion	177
6	Conclusion générale et perspectives	180
6.1	Contributions	181
6.1.1	L'arrière-plan comme objet sonore	181
6.1.2	Réalisation d'un synthétiseur d'arrière-plans so- nores urbains dynamiques	182
6.2	Analyse critique et travaux futurs	183
6.2.1	Développements du synthétiseur	183
6.2.1.1	Différentes options de fenêtrage des grains	184
6.2.1.2	<i>Warping</i> audio	184
6.2.1.3	Mise en espace	185
6.2.1.4	Nouvelles sources sonores	185
6.2.2	Validations expérimentales	186
6.2.3	Intégrations futures	187
A	Sonagrammes	188
B	Exemples de l'utilisation du synthétiseur	199
B.0.4	Ajout de nouveaux fichiers audio	199
B.0.5	Composition d'une ambiance	200
B.0.5.1	Recréer des plans sonores	201

B.0.5.2	Mixage dynamique de différentes am- biances	201
B.0.6	Courbes d'automation	202
C	Expérimentation : choix des échantillons et production des fichiers sonores	205
C.1	Production des resynthèses avec <i>timbreID</i>	206
C.2	Production des resynthèses avec <i>Grainer</i>	211
D	Questionnaire pour la série A d'échantillons	213
E	Résultats de l'expérimentation	215
F	Résultats des tests t de Student	216

Table des figures

2.1	L'écologie sonore comme étude des interactions entre les individus et leur environnement par le biais du son, <i>in</i> [Wri00]	26
2.2	Fondements intellectuels de l'écologie acoustique paysagère, <i>in</i> [PFG ⁺ 11]	27
2.3	Schéma conceptuel décrivant les processus sous-jacents du paysage sonore. <i>in</i> [PFG ⁺ 11]	27
2.4	Cycle annuel du paysage sonore naturel de la côte Ouest de la Colombie-Britannique (Canada) par volume relatif des sons <i>in</i> [Tru78]	30
2.5	Exemples de sonagrammes annotés comparant quatre enregistrements d'un même paysage à des heures différentes <i>in</i> [PFG ⁺ 11]	31
2.6	Exemple de représentation d'un espace de timbre tridimensionnel extrait de l'analyse de proximité des jugements de similarité <i>in</i> [Ver03]	37
2.7	Fluctuations de l'amplitude (gauche) et de la tonalité (droite) pour les paysages sonores naturels (N3 et N4), ruraux (R8 et R11) et urbains (U9 et U11), <i>in</i> [BDD06]	41
2.8	Mélange des champs acoustiques direct et indirect issus d'une source sonore et formant l'ensemble perçu au point d'écoute	42
2.9	Représentation de l'effet canyon ainsi que de la dispersion acoustique d'une source sur les façades, <i>in</i> [IO05]	42
2.10	Comparaison de la répartition des fréquences entre les arrière-plans sonores urbains enregistrés (points noirs) et simulés (points blancs), présentée dans [Lyo74] selon le modèle proposé dans [SO72]	44
2.11	Simplification de la taxonomie de Gaver, <i>in</i> [VAKMP10]	48

2.12	Comparaison entre les descriptions verbales de paysage sonore urbain fournies par ses habitants et par des planificateurs urbains <i>in</i> [RD05]	50
2.13	Image extraite du film <i>Berlin, symphony of a great city</i> (1927)	56
2.14	Image extraite du jeu vidéo <i>Bioshock</i> (2K Games, 2007), vue du principal environnement du jeu : la ville <i>Rapture</i>	64
2.15	Image extraite du jeu vidéo <i>Prototype</i> (Radical Entertainment, 2009)	66
2.16	Photographie du panneau avant du module de synthèse par table d'ondes <i>E350 Morphing Terrarium</i> (Synthesis Technology, 2012), ce module numérique est destiné aux synthétiseur modulaires de format <i>Eurorack</i>	76
2.17	Photographie d'un système PPG <i>Wave 2</i> complet (1981) incluant le moniteur <i>Waveterm</i> de visualisation de tables d'ondes.	77
2.18	Le bruit et les textures sonores montrant des caractéristiques constantes sur le long-terme, <i>in</i> [SAP97]	90
3.1	Représentation de l'intensité dans le temps d'une réponse impulsionnelle	107
3.2	Relations entre les niveaux sonores des champs direct et diffus	107
4.1	Architecture simplifiée du prototype de synthétiseur : de gauche à droite, les oscillateurs composés de quatre <i>Grainers</i> dont le mélange et le filtrage sont contrôlés par le <i>Blender</i> , les oscillateurs sont ensuite mélangés dans le Mixeur principal.	134
4.2	Détails du <i>patch Grainer</i>	138
4.3	Détails du <i>Grainer</i> , ici le sous- <i>patch</i> servant à calculer le ratio entre la taille de grain et la taille des fondus.	139
4.4	<i>Blender</i> permettant de parcourir les différents <i>Grainers</i> d'un oscillateur	141
4.5	Répartition des paramètres du synthétiseur sur un contrôleur Korg nanoKONTROL2	149
4.6	Enregistreur portable Olympus LS-5	150
5.1	Schéma d'ensemble de la conception du projet Terra Dynamica	158

5.2	Sonorisation de Terra Dynamica : schématisation de l'articulation dynamique entre Fmod (sources individuelles) et le synthétiseur réalisé dans Pure Data (arrière-plans), présentation du type de données transmises de la simulation vers Fmod et Pure Data	161
5.3	Fmod : configuration des paramètres 3D d'une source sonore.	162
5.4	Fmod : fenêtre <i>Engine Designer</i>	163
5.5	Terra Dynamica : rendu visuel (1/2)	164
5.6	Terra Dynamica : rendu visuel (2/2)	166
5.7	Terradynamac~ : détail des sorties de l'objet.	167
5.8	Pure Data : <i>patch</i> utilisé comme interface de test pour les passations	175
5.9	Moyennes et écart-types des résultats de l'expérimentation.	175
A.1	Sonagramme, promenade sonore de la station Quatre Septembre à Notre Dame de Lorette, Paris	189
A.2	Sonagramme, ambiance sonore de la ville de Liège 1	190
A.3	Sonagramme, ambiance sonore de la ville de Liège 2	191
A.4	Sonagramme, ambiance sonore de la ville de Liège 3	192
A.5	Sonagramme, ambiance sonore de la ville de Liège 4	193
A.6	Sonagramme, ambiance du site BNF captée depuis le parvis de la bibliothèque, Paris	194
A.7	Sonagramme, ambiance du site BNF captée à proximité du périphérique, Paris	195
A.8	Sonagramme, San Francisco, ambiance captée au niveau de la circulation en centre ville	196
A.9	Sonagramme, San Francisco, ambiance captée depuis le huitième étage d'un parking en centre ville	197
A.10	Sonagramme, San Francisco, ambiance captée depuis les hauteurs du parc Buena Vista, légèrement excentré du centre ville	198
B.1	Exemples de courbes d'automation du synthétiseur d'ambiances urbaines	203
C.1	Sonagramme de l'échantillon A de l'expérimentation	205
C.2	Sonagramme de l'échantillon B de l'expérimentation	206
C.3	Sonagramme de l'échantillon C de l'expérimentation	206

C.4	Sonagramme de l'échantillon D de l'expérimentation . .	207
C.5	Pure Data, <i>timbreID</i> : fenêtres de configuration (bas) et de projection de timbre (haut) du <i>patch</i> « <i>timbre-space</i> ».	209
C.6	Pure Data : <i>patch</i> utilisé pour générer les fichiers de resynthèse granulaire	211

Chapitre 1

Introduction générale

Sommaire

1.1	Contexte de la thèse : le projet Terra Dynamica	18
1.2	Méthodologie	19
1.3	Plan de la thèse	20

L'évolution des mondes virtuels va dans le sens d'une économie du temps de développement, d'une économie des ressources techniques disponibles, tandis que la taille des univers augmente ainsi que leur richesse et leur diversité. L'implémentation de méthodes procédurales, c'est-à-dire de processus programmés pour générer du contenu selon des règles prédéfinies, accompagne cette tendance avec pour objectif de simplifier le développement tout en nourrissant la diversité intrinsèque de l'environnement virtuel. Les méthodes procédurales sont potentiellement des sources de variations infinies en réponse aux changements d'état du monde virtuel.

Le projet Terra Dynamica¹ est un exemple d'environnement virtuel complexe dont la pluralité des applications requiert une grande adaptabilité logicielle. Les différentes composantes du projet (animations, intelligence artificielle, rendu sonore...) doivent être conçues de manière à répondre à cette pluralité.

Au cours de la dernière décennie, les méthodes procédurales ont principalement été développées sur le plan graphique, notamment pour les animations de personnages dans les jeux vidéo [AF02]. Des stra-

1. www.terradynamica.com

tégies équivalentes commencent à se développer pour la production sonore dédiée aux environnements interactifs [Far07a], ces démarches restent encore minoritaires. La production sonore de la majorité des environnements virtuels consiste en un mixage dynamique d'échantillons traités avec quelques effets temps réel, cette démarche efficace ne permet cependant qu'une palette limitée de nuances dans le rendu. L'audio procédural consiste à produire des rendus non-linéaires en temps réel selon un jeu de règles programmées et des flux de variables entrantes ([Ven09] p.111). Si certains domaines sont plus difficiles à appréhender de façon procédurale, comme la synthèse de la voix, d'autres font l'objet d'applications fructueuses comme la musique générative. Ces mêmes recherches sur les ambiances et les sons environnementaux sont en plein essor et de nombreuses solutions sont à explorer ou à optimiser pour des applications temps réel. La faculté des pistes sonores d'ambiance à caractériser un contexte en font des éléments très présents dans l'écriture sonore des mondes virtuels. Le travail des designers sonores assignés à cette tâche répond à deux principales contraintes : la clarté du message transmis par l'ambiance (caractérisation d'un lieu, d'un état de l'environnement, charge émotionnelle associée) et la sensation de renouvellement constant de ce message (la sensation de répétition pouvant briser le caractère immersif).

Ce travail de recherche s'inscrit dans le cadre particulier de la ville virtuelle. La sonorisation des ambiances urbaines pose des enjeux spécifiques liés à l'acoustique de la ville, à la densité et au type de sources sonores qui la peuplent. D'autre part, l'interactivité de la ville virtuelle soumet les ambiances sonores à l'influence de deux principaux flux de paramètres temps réel : les fluctuations de l'état de la ville (climat, variations du trafic selon l'heure de la journée...) et la liberté de déplacement de l'écouter. La conception des ambiances doit tenir compte de l'interactivité et des contraintes temps réel pour créer et maintenir la sensation d'immersion dans l'environnement virtuel.

Il s'agit, dans un premier temps, de définir les principales composantes du son urbain et plus particulièrement le rôle des arrière-plans sonores. La notion d'arrière plan sonore urbain a-t-elle un sens théorique et surtout pratique dans des systèmes de type jeu vidéo ? Peut-on, à partir de cette analyse, concevoir des stratégies de sonification utilisant les arrière-plans comme objets sonores pour l'écriture de scènes virtuelles interactives ? L'enjeu serait ici de donner au concepteur sonore les moyens de gérer dynamiquement l'importance de l'arrière-plan

par rapport aux autres objets sonores de la scène, ceci en tenant compte des paramètres temps réel fournis par le moteur de jeu (par exemple la densité du trafic) et des variations de la position de la caméra et du point d'écoute dans un système interactif. Nous proposons donc le développement d'un outil pratique de synthèse sonore, interactif et temps réel, dédié aux arrière-plans urbains. Enfin, nous intégrons le synthétiseur développé dans une stratégie de sonification de scènes urbaines complexes avec l'objectif d'un rendu perceptivement crédible pour des applications de type jeu vidéo.

Cette thèse se situe dans le prolongement des recherches en audio pour les médias interactifs effectuées au laboratoire Cédric ([Gon08], [Ven09], [NLP07]). Ces recherches mettent en avant les problématiques liées à l'écriture sonore interactive mais également aux technologies mises en œuvre pour faire vivre cette écriture.

La spatialisation sonore en 3D, la synthèse temps-réel ou encore l'optimisation du rendu font partie intégrante des enjeux de la création d'un média audiovisuel interactif. La composition musicale et le design sonore doivent également prendre en compte la non-linéarité et le caractère dynamique, inhérents à une bande-son interactive.

Ces besoins spécifiques ont fait émerger des outils dédiés permettant d'intégrer les sources sonores, de spécifier leurs comportements, de leur donner une position dans l'espace de la scène virtuelle et de lier un certain nombre de paramètres aux événements du média sonorisé.

Ces dernières années, les outils dédiés au rendu graphique intégrant des fonctionnalités temps-réel complexes, que ce soit pour des animations procédurales ou des niveaux de détails dynamiques, se sont développés. La partie sonore ne fait encore que peu appel à ce genre de techniques. Les outils sonores proposés aujourd'hui héritent pour bonne part des méthodologies utilisées pour le cinéma, c'est-à-dire l'enregistrement d'une source, suivi de son édition et enfin sa superposition à l'image lui correspondant.

Si des outils de traitement du signal en temps réel sont de plus en plus intégrés à la création de bande-sons interactives, les techniques procédurales en matière de composition ou de synthèse sonore restent en retrait. Des solutions allant dans ce sens existent pourtant et leur intégration, à plus grande échelle et plus en amont dans la chaîne de production, suscite de vifs intérêts de la part de certains développeurs ou artistes.

L'utilisation de ces technologies s'intercale entre deux enjeux. D'un côté, elles sont une solution pour utiliser moins de ressource mémoire puisque toute source synthétisée n'a pas besoin d'être présente sous la forme d'un fichier audio dans les sources du média. D'un autre côté, la synthèse sonore (et les technologies procédurales dans leur ensemble) peut être consommatrice de davantage de ressources processeur. Un compromis entre espace mémoire et temps de calcul ainsi qu'une étape d'optimisation sont donc nécessaires. Lors de la création d'un média, l'ensemble des ressources disponibles est réparti entre les différentes parties qui les consomment, telles que les rendus graphiques ou les système d'intelligence artificielle. De manière générale, lors de la conception d'un jeu vidéo, la partie sonore se voit attribuer aux alentours de cinq pour cent de l'ensemble des ressources disponibles. L'optimisation est un point crucial du développement dans la mesure où elle conditionne en partie les étapes de la production sonore.

1.1 Contexte de la thèse : le projet Terra Dynamica

Ce travail de thèse est étroitement lié au développement du projet Terra Dynamica². Terra Dynamica est la continuité des projets Terra X antérieurs (Terra Magna, Terra Numerica) visant la collecte et la numérisation de données urbaines (trafic, flux, géométrie et apparence des façades, comportements). Le projet Terra Dynamica avait pour objectif d'animer ces données en recréant, en 3D, l'environnement de la ville de Paris et en le peuplant de milliers d'acteurs selon différents scénarios. Le rendu devait intégrer, en temps réel, de la gestion de foules, de trafic et d'intelligence artificielle.

La partie sonorisation de Terra Dynamica incombait au Cnam sous la forme du développement d'un moteur sonore 3D. L'objectif de la sonorisation était double, d'une part renforcer le caractère immersif de la simulation par l'ajout d'une trame sonore crédible, et d'autre part se servir du son comme d'un média porteur d'informations qui ne peuvent être dispensées par le visuel uniquement (par exemple, la sonorisation d'événements situés hors-champs). Le principal enjeu de la

2. Terra Dynamica est un projet labellisé par les pôles de compétitivité Cap Digital et Advancity, et qui associe une grande entreprise, Thales (porteur du projet), et 4 laboratoires de recherches universitaires et 4 PME. www.terradynamica.com

sonorisation de Terra Dynamica concerne la densité des scènes urbaines au regard du nombre d'acteurs les peuplant. Le travail de thèse que nous présentons propose une stratégie perceptive, intégrée à la version finale de Terra Dynamica, permettant de résoudre cet enjeu.

1.2 Méthodologie

Le calcul du rendu acoustique, en temps réel, de centaines ou milliers de sources en simultané est une démarche qui peut s'avérer excessivement coûteuse en temps de calcul [TGD04]. Les limites induites par le modèle perceptif humain [Bre94] compromettent l'intérêt de maintenir l'individualité de la sonorisation des sources au-delà d'une certaine distance par rapport au point d'écoute, relativement à la densité de la scène.

D'autre part, les stratégies de sonorisation utilisées dans la production au cinéma ou pour le jeu vidéo montrent l'importance de conserver une lisibilité de la scène sonore ainsi qu'une forme de cohérence par rapport au visuel. Dans le cadre de la sonorisation d'une simulation urbaine, il convient donc de trouver un équilibre entre le réalisme et le sens de la trame sonore.

Ce travail de thèse, en se fondant sur la théorie du paysage sonore, les méthodes d'écriture pour l'audiovisuel interactif et la recherche en synthèse de textures sonores, envisage le rendu de scènes urbaines complexes comme la superposition de différentes strates sonores recomposant la perspective acoustique et distinguant les sources selon leur description sémantique.

C'est de cette partition qu'émerge le besoin de créer un outil dédié aux arrière-plans sonores, objets ubiquitaires, englobant l'intégralité d'une scène, par opposition aux sources individuelles. L'arrière-plan n'est pas l'illustration d'une action en cours et visible, aucune des sources qui le compose n'est clairement identifiable, son rôle, en tant qu'objet, est davantage descriptif. Cette couche sonore permet de rendre compte, sur un plan perceptif, de l'état de la scène. Du point de vue de la composition et du mixage, l'arrière-plan est un drone ou un cluster servant de support aux sources individuelles et aux groupes de sources sonorisées qui composent la scène, c'est un liant qui renforce la perspective acoustique de la scène.

Sur la base d'enregistrements bruts, la sonorisation d'un média peut nécessiter de créer des boucles audio de plusieurs minutes. Un tel procédé consomme rapidement d'importantes ressources mémoire et crée le risque d'une sensation de répétition. L'un des enjeux du travail de thèse présenté ici est d'adapter les méthodologies de la synthèse de texture afin de produire des flux audio d'une durée plus importante que celle des échantillons utilisés au départ. Nous émettons l'hypothèse que la nature sonore des arrière-plans urbains se prête particulièrement à l'utilisation de technologies procédurales. Nous proposons donc un outil interactif de synthèse dédié aux ambiances sonores urbains et dont le rendu cherche la crédibilité perceptive. Les domaines d'application de cet outils sont principalement liés à la production d'ambiances sonores dans les jeux vidéo et plus largement dans les environnements virtuels mais sans pour autant exclure la sonorisation de média linéaires tels que le cinéma.

Cette thèse a donc pour objectif de définir, spécifier et prototyper un outil de synthèse de l'arrière-plan urbain utilisable dans des applications fortement interactives telles que le jeu vidéo.

1.3 Plan de la thèse

Comment l'articulation acoustique de la ville est-elle comprise? Comment son pendant virtuel est-il construit? Ces deux questions mettent en regard la notion d'arrière-plan sonore urbain du point de vue réel et du point de vue virtuel et, plus globalement, interrogent la compréhension de la dynamique du son urbain.

Nous procédons dans un premier temps à un état de l'art du son urbain et sa mise en espace (Chapitre 2). Nous y abordons transversalement l'architecture et les théories d'analyse du paysage sonore ainsi que les particularités de l'acoustique urbaine qui font émerger un fond sonore diffus.

Nous décrivons ensuite les méthodes d'écriture du son dans la ville, selon quatre axes. Le premier concerne les plans d'aménagements urbains qui constituent un trait d'union entre la ville telle que présentée précédemment et sa virtualisation. Le deuxième axe concerne l'écriture sonore pour les représentations audiovisuelles linéaires du cinéma et des arts numériques. Le troisième axe décrit les représentations cartographiées en incluant les cartes sonores participatives. Le quatrième

axe dresse un bilan des techniques utilisées pour sonoriser les villes virtuelles des médias interactifs et des jeux vidéo. Ces axes nous aident à cerner la logique d'écriture propre à la virtualisation de la ville ainsi que les contraintes qui y sont liées en matière de gestion de l'espace et de la spatialisation.

Une fois ces processus dynamiques décrits, un certain nombre de problématiques liées à la synthèse sonore d'un environnement urbain complexe se dégagent (Chapitre 3), parmi lesquelles nous choisissons de nous intéresser plus spécifiquement aux arrière-plans sonores sans toutefois négliger les contraintes de niveaux de détails ainsi que les aspects sémantiques. Après un comparatif des différentes méthodes de synthèse sonores qui peuvent être mises en œuvre, nous définissons plus précisément la notion d'arrière-plan urbain au travers de ses particularités spectrales et dynamiques.

Le chapitre suivant (Chapitre 4) présente les étapes de développement du prototype de synthétiseur temps-réel dédié aux ambiances urbaines réalisé dans le cadre de cette thèse. Nous décrivons ici les expérimentations préliminaires à la réalisation du synthétiseur, le choix de Pure Data comme environnement de développement ainsi que les aspects techniques du prototype.

Nous présentons ensuite (Chapitre 5) l'implémentation du prototype dans le projet TerraDynamica et sa validation par le biais d'écoutes comparatives.

Le dernier chapitre (Chapitre 6) apporte une conclusion à l'ensemble de ces travaux ainsi qu'une approche critique de la méthode et des résultats obtenus tout en proposant des développements futurs.

Chapitre 2

État de l’art : Espaces sonores urbains

Sommaire

2.1	Concepts fondamentaux	24
2.1.1	Paysages, écologie, environnements, scènes sonores	25
2.1.2	Timbres sonores	36
2.2	La ville sonore	39
2.2.1	Analyse acoustique	40
2.2.2	Analyse perceptive	45
2.3	Virtualisation et sonorisation de la ville . .	53
2.3.1	Villes sonores au cinéma	53
2.3.2	Villes sonores interactives	59
2.4	Outils de sonorisation des villes virtuelles .	72
2.4.1	Synthèse en temps réel d’ambiances urbaines	72
2.4.2	Outils dédiés au jeu vidéo	82
2.5	L’arrière-plan urbain comme texture sonore	89
2.5.1	Définition d’une texture sonore	89
2.5.2	Synthèse de textures	91
2.5.3	Arrière-plans sonores : textures complexes .	94
2.6	Conclusion	96

Cette thèse a pour objet de définir, spécifier et prototyper un outil de synthèse de l'arrière-plan urbain, utilisable dans des applications fortement interactives telles que le jeu vidéo. L'objectif de ce chapitre est de donner une définition aussi complète que possible aux différents termes de notre sujet et de mettre en évidence les travaux de recherche, les outils et les pratiques qui démontrent la nécessité et la faisabilité de ce travail.

Dans un premier temps, nous précisons la notion d'arrière-plan urbain telle qu'il est possible de la concevoir dans les villes virtuelles. Pour cela, nous nous appuyons sur des concepts fondamentaux liés au paysage, à l'écologie et aux environnements sonores. Ceci nous amène à préciser ce qu'est la ville sonore, d'un point de vue acoustique et d'un point de vue perceptif. Nous examinons ensuite comment cette ville sonore est traitée dans des fictions linéaires et des fictions interactives.

La seconde partie de cet état de l'art s'intéresse aux concepts et aux outils existants pour traiter la notion d'arrière-plan sonore urbain dans le cadre de fictions interactives. Nous montrons que l'arrière-plan urbain peut être considéré comme une texture sonore et synthétisé en tant que tel. Nous examinons ensuite les différentes méthodes de synthèse qui s'offrent à nous, nous rappelons les contraintes liées au type d'applications interactives que nous visons et présentons brièvement les principales applications de sonorisation des jeux vidéo.

2.1 Concepts fondamentaux

Cette section aborde les définitions des concepts clefs relatifs à la thèse. Il s'agit de concepts aidant à cerner la nature des phénomènes sonores que nous abordons dans ce travail de thèse et leur donnant un contexte théorique. Les phénomènes sonores de la ville sont abordés ici selon deux axes principaux : un axe définissant le sonore comme une organisation de sons liée à un contexte environnemental et un axe définissant le sonore par des caractéristiques timbrales. La compréhension et l'analyse perceptives de la ville sonore servent, plus loin, de support à la proposition d'un outil de synthèse sonore environnementale dédié aux villes virtuelles ainsi que de stratégies de sonification dans lesquelles viennent s'intégrer le synthétiseur décrit.

Parmi les études proposant des analyses de la ville sonore, nous choisissons de nous intéresser principalement à celles inspirées de la théorie du *paysage sonore*, ce choix est motivé par plusieurs raisons. La théorie du paysage sonore propose des méthodologies d'analyse de l'environnement réel et de création appliquée à des installations, des œuvres audiovisuelles et des environnements virtuels. Nous souhaitons nous inspirer de ces méthodologies la conception et l'intégration d'un outil de création et de design sonore. La théorie du paysage sonore constitue une approche environnementale c'est-à-dire que le phénomène sonore est considéré en lien avec l'environnement dans lequel il se produit. Cette approche offre un cadre d'interactions possibles entre un outil de synthèse sonore et l'environnement virtuel dans lequel il s'intègre.

Nous associons ensuite la compréhension des phénomènes urbains à l'analyse des timbres qu'ils donnent à entendre. La recherche sur la perception et la classification des timbres nous apparaît importante dans ce qu'elle peut apporter comme informations sur le ressenti de l'environnement sonore ainsi que des principaux paramètres qui permettent de définir la couleur d'un son. Une étude reposant sur la notion de timbre ne distingue pas la hauteur ou la durée des sons mais se focalise sur des attributs caractérisant la matière sonore. Ces études furent d'abord menées sur des sons musicaux, nous verrons dans quelle mesure elles sont transposables à des sons environnementaux.

2.1.1 Paysages, écologie, environnements, scènes sonores

En 1977, Raymond Murray Schafer [MS10] définit la notion de paysage sonore (*soundscape*) :

The soundscape is any acoustic field of study. We can isolate an acoustic environment as a field of study just as we can study the characteristics of a given landscape. However, it is less easy to formulate an exact impression of a soundscape than of a landscape (p. 7)

Pour Bernie Krause, dans une perspective beaucoup plus large, le paysage sonore regroupe tous les sons présents dans un environnement à un moment donné [PFG⁺11]. Pour Barry Truax [Tru78], la notion de paysage sonore peut faire référence autant à un environnement réel qu'à des constructions abstraites ou des environnements artificiels. Elle n'est pas simplement un synonyme de l'environnement acoustique, elle fait référence à la manière dont les individus et la société dans son ensemble comprennent l'environnement acoustique, comme support de communication, au travers de l'écoute [Tru84].

Ces quelques définitions font ressortir que la perception de ce qui constitue un paysage sonore est en partie relative au contexte d'écoute et aux informations que l'écouter cherche à recueillir. Ainsi, Barry Truax définit le paysage sonore comme un environnement dont l'analyse permet de faire ressortir la manière dont il est perçu et compris par les individus ou la société [Tru78].

D'un point de vue qualitatif, un paysage *hi-fi* est distingué d'un paysage *lo-fi* [Tru08] en rapport à la clarté de sa perspective acoustique, c'est-à-dire de l'intelligibilité des éléments sonores les plus distants du point d'écoute du paysage. Le paysage sonore urbain est souvent pris comme exemple caractéristique d'un paysage *lo-fi* avec une perspective acoustique réduite et une dynamique très faible. Les sources sonores peuplant un paysage *hi-fi* sont variées et fortement localisées ; au contraire, un paysage *lo-fi* est uniforme dans son amplitude et sa répartition spatiale. Le paysage *lo-fi* est souvent caractérisé par l'émergence de sources sonores issues de l'industrialisation qui masquent la diversité naturelle, c'est la forme dominante de paysage dans les environnements urbains.

Dans *Le Paysage sonore*, Murray Schafer introduit également le concept d'*écologie sonore* qui consiste en l'étude de la relation entre les organismes vivants et leur environnement sonore (Fig. 2.1). Ce concept est complété par celui d'*écologie acoustique* [Wri00] qui propose de comprendre l'environnement acoustique comme une composition musicale dont nous sommes en partie responsables du fait de nos actions dans l'environnement.

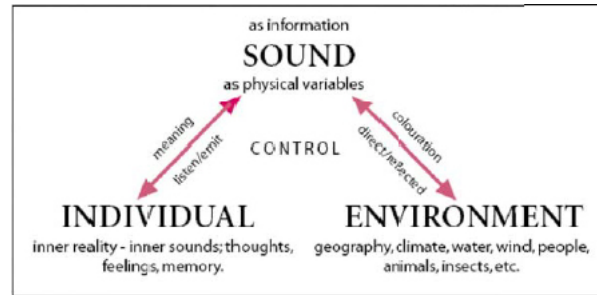


FIGURE 2.1 – L’écologie sonore comme étude des interactions entre les individus et leur environnement par le biais du son, *in* [Wri00]

L’émergence et la croissance des notions de paysage et d’écologie sonore mènent à la diversification des axes de recherches ainsi qu’à leur spécialisation pour tel ou tel domaine. Dans un souci d’unification, Pijanowski et al. [PFG⁺11] proposent en 2011 ce schéma (Fig. 2.2) afin d’explicitier les relations entre les différentes branches de recherche qui gravitent autour de ces notions.

La Fig. 2.2 illustre la diversité des champs de recherche qui participent à l’écologie sonore, du point de vue de la notion de paysage. Ce schéma est complété par la Fig. 2.3 qui montre comment les facteurs naturels et humains influencent les *patterns* acoustiques sociaux et écologiques.

Les points d’entrée proposés par la Fig. 2.2 ne sont pas exhaustifs, il s’agit principalement, pour les auteurs, de mettre en évidence l’impact potentiel d’une multitude de facteurs sur la dynamique sonore d’un paysage. L’interaction de ces facteurs peut être unique à chaque paysage et la classification des paysages sonores fait l’objet d’études à part entière [PBA⁺08].

Parmi les champs de recherche mis en avant, l’étude de l’acoustique urbaine tient une place singulière. Il s’agit d’étudier les agrégats sonores qui s’entrelacent dans l’environnement bâti et qui émergent

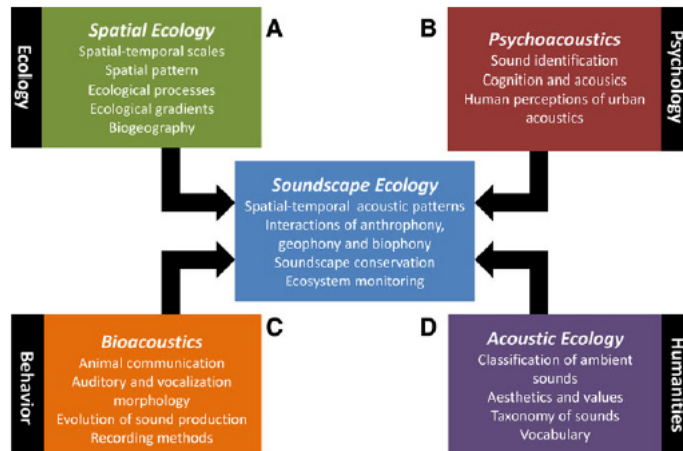


FIGURE 2.2 – Fondements intellectuels de l'écologie acoustique paysagère, *in* [PFG⁺11]

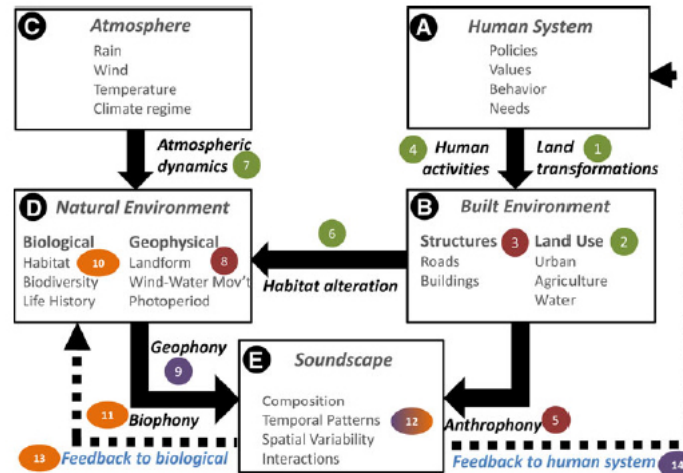


FIGURE 2.3 – Schéma conceptuel décrivant les processus sous-jacents du paysage sonore. *in* [PFG⁺11]

naturellement en raison des activités conduites dans l'espace public [BDD06]. En raison de l'impact des propriétés acoustiques sur la qualité de vie des résidents, la recherche sur le paysage sonore urbain peut s'étendre à la planification urbaine, la modélisation de l'acoustique ainsi que la simulation de la propagation sonore qui ont fait l'objet

de nombreuses recherches depuis des années¹. L'étude de l'environnement bâti se fait également par le biais d'évaluations subjectives afin de déterminer comment les habitants perçoivent leur environnement².

Nous n'entrerons pas ici dans une étude détaillée des diverses ramifications de recherche et de pratique liées à la notion de paysage sonore mais traiterons en priorité les concepts et études qui permettent d'appréhender l'espace sonore urbain, autant du point de vue théorique qu'applicatif.

2.1.1.1 Paysage et écologie sonores : des outils d'analyse

Les théories du paysage et de l'écologie sonores fournissent des outils d'analyse de l'environnement ou de scène sonores. Elles ont en commun de proposer une segmentation et une structuration de ce qui est audible, d'un point de vue physique, perceptif ou bien encore sémantique.

Dans sa définition la plus communément admise, le paysage sonore divise le contenu sonore de tout environnement en trois sous-ensembles : arrière-plan (ou drone, *keynote*), signaux sonores (*sound signals*) et repères sonores (*soundmarks*). Ces trois catégories, proposées par Murray Schafer, remplissent chacune un rôle différent dans la perception de l'environnement.

L'arrière-plan est défini par son ubiquité spatiale [Wri00] et sa relative constance dans le temps. Cette strate sonore caractérise un environnement et sert de support aux autres sons qui le composent. Murray Schafer emploie à dessein le terme de *keynote*, soit *note fondamentale*, pour désigner ce qui marque la tonalité d'un paysage. Les repères sonores (*soundmarks*) sont caractéristiques du lieu pour celui qui les entend, ils servent de balise sonore afin d'identifier un lieu. Il peut s'agir du son de Big Ben, des sirènes de New-York ou de tout son familier et/ou unique pour l'auditeur. Les signaux sonores (*sound signals*) regroupent l'ensemble des sons qui, n'entrant pas dans les deux catégories précédentes, illustrent les activités ou la géographie du paysage [Fin09]. La théorie du paysage sonore propose ainsi une première approche sémantique d'un environnement et permet d'avoir des repères d'analyse en segmentant l'espace perçu en deux plans distincts (pre-

1. Voir Chapitre 2, Section 2.1

2. Voir Chapitre 2, Section 2.2

mier plan et arrière-plan) [SJB14].

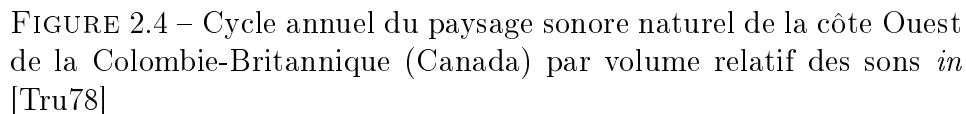
Dans une perspective écologique [PVRD⁺11], le paysage se subdivise également en trois phonies distinctes : biophonie, géophonie et anthropophonie [VRPDP11].

La biophonie regroupe les sons créés par des organismes biologiques (insectes, amphibiens, oiseaux, mammifères). Ce sont des signaux porteurs d'informations et donc complexes. À ce sujet, Bernie Krause propose l'hypothèse de la « niche acoustique » : le paysage sonore serait la résultante d'une dispute de ressources autour de l'espace spectral sous la forme d'occupation de bandes de fréquences ([Kra87] et [Kra02]). La géophonie est l'ensemble des sons produits par les mouvements de l'air et de l'eau (cours d'eau, pluies, vents), ceux-ci varient donc souvent avec le climat. L'anthropophonie regroupe les sons émis par des objets fabriqués par les êtres humains tels que les machines, les sons de roulement, les cloches, les sirènes.

Le concept de paysage sonore est un référent important au cours de ce travail car il est un outil de compréhension des représentations réelles et virtuelles.

Parmi les outils d'analyse dérivés du concept de paysage sonore, notons les *fields recordings* et les *soundwalks* (promenades sonores). Le *field recording* désigne de manière générale l'ensemble des pratiques d'enregistrement sonore se déroulant en dehors du cadre d'un studio dédié ; avec l'émergence de l'écologie sonore, cette pratique s'est accrue dans un but de documentation et d'analyse des paysages et comme un moyen de récupérer des matériaux sonores sans un processus de création. Barry Truax définit les promenades sonores comme une forme de participation active à l'écoute d'un paysage sonore ; les variations en sont nombreuses mais le but essentiel reste d'encourager les participants à une écoute détaillée ainsi que d'émettre des jugements critiques à propos des sons entendus et de leur impact qualitatif sur l'équilibre de l'environnement [Tru78]. Plus que des moyens d'observation, les promenades sonores peuvent devenir des outils d'analyse du paysage. Introduites par Murray Schafer en même temps qu'émerge la notion de paysage sonore, elles sont une méthode empirique pour identifier un paysage et ses composantes en plusieurs endroits. Le concept a par la suite été adapté afin de correspondre à des besoin d'observation et d'évaluation précis variants selon les environnements [ABD⁺08].

La perception subjective est au centre de l'analyse du paysage sonore, la principale méthode de conservation d'un paysage en est l'enregistrement. Cependant, des référents visuels peuvent venir en support, autant pour garder trace des observations que pour leur donner une structure à même d'en multiplier les usages.



30

temporelles (Fig. 2.4), référençant des niveaux ou des sources sonores et n'excluant pas des traitements statistiques. Ces outils d'analyse permettent d'effectuer des comparaisons temporelles ou de croiser les données avec d'autres disciplines comme la géographie, la sociologie ou l'ethnographie.

Parmi ces outils, les représentations temps-fréquences appelées sonagramme proposent certaines singularités. Un sonagramme est obtenu sur la base d'une transformée de Fourier, il représente trois dimensions sonores. Le temps est sur l'axe des abscisses et les fréquences sont sur l'axe des ordonnées. L'énergie attribuée à une fréquence à un instant donné est représentée par un code couleur ou une nuance de teinte. La lecture d'un sonagramme (parfois aussi appelé spectrogramme) se fait à la manière d'une partition : les hauteurs sont disposées linéairement avec les fréquences élevées en haut de la portée [PvRD⁺11]. Le sonagramme permet d'afficher une vue d'ensemble des différentes intensités sonores d'un fichier [VRPDP11] et, de fait, de créer des points de comparaison entre différents enregistrements. Comme présentés par

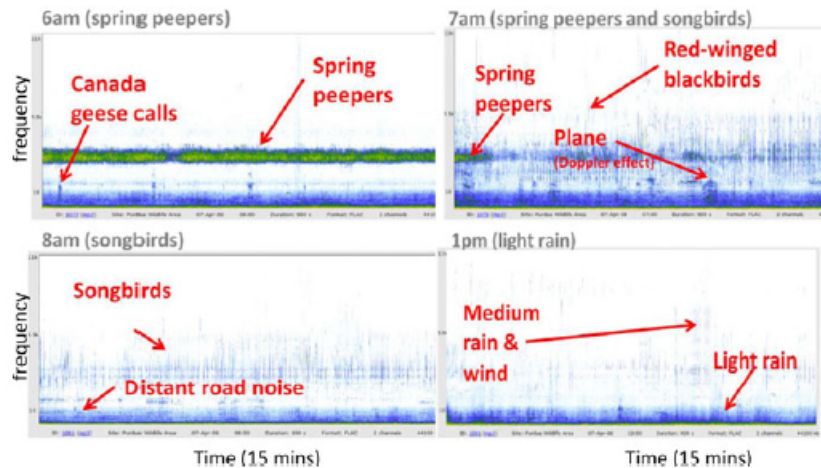


FIGURE 2.5 – Exemples de sonagrammes annotés comparant quatre enregistrements d'un même paysage à des heures différentes *in* [PFG⁺11]

la Fig. 2.5, les sonagrammes sont souvent annotés afin de référencer des sources ou des événements de l'enregistrement [PFG⁺11]. Développé au GRM, le logiciel Acousmographie³ offre ainsi des outils d'annotation

3. <http://www.inagrm.com/accueil/outils/acousmographie>

intégrés. Si les sonagrammes sont historiquement utilisés pour analyser des caractéristiques spectrales subtiles sur des enregistrements de courte durée (généralement quelques dizaines de minutes), les enjeux d'analyse du paysage sonore couvrent des changements se déroulant sur des durées plus importantes. Ainsi, en 2009, Kaliski et Joyce [KJ09] proposent un outil pour réaliser des sonagrammes couvrant une période de deux semaines à partir de micro-enregistrements.

2.1.1.2 Environnements sonores

Un autre concept important précédemment mentionné est celui d'environnement sonore. Une définition contemporaine de celle du paysage sonore, proposée par Vanderveer (1979) (et reprise par Misdariis dans [MMS⁺10]) présente les sons environnementaux comme tout événement acoustique audible causé par des déplacements dans l'environnement humain quotidien. Des événements réels en sont à l'origine, ils sont d'ordinaire plus complexes que des sons artificiellement reproduits à l'aide de sinusoïdes, ils sont porteurs de sens dans la mesure où ils caractérisent les événements de l'environnement. Les sons environnementaux ne font pas partie d'un système de communication et ne sont pas des sons de communication. Ils sont à considérer dans leur sens littéral et non par une interprétation symbolique⁴.

Le concept de paysage sonore semble englober celui d'environnement sonore, le paysage pouvant inclure des constructions sonores abstraites. La notion d'environnement insiste plus spécifiquement sur les informations et le sens rapportés d'événements réels. Les sons environnementaux, tout comme le paysage, englobent des interactions et des timbres complexes, la notion d'environnement privilégie cependant les contextes incluant la présence et l'action humaines. En pratique, les deux termes sont souvent confondus. Il est tout de même possible de reconnaître une dimension analytique et créative à la notion de paysage sonore tandis que le concept d'environnement sonore va plus volontiers

4. Définition originale : «... *any possible audible acoustic event which is caused by motions in the ordinary human environment. (...) Besides 1) having real events as their sources (...), 2) [they] are usually more 'complex' than laboratory sinusoïds, (...), 3) [they] are meaningful, in the sense that they specify events in the environment. (...), 4) the sounds to be considered are not part of a communication system, or communication sounds, they are taken in their literal rather than signal or symbolic interpretation*» in [MMS⁺10], d'après Vanderveer (1979).

faire référence à des ensembles complexes de sons répartis dans l'espace.

2.1.1.3 Le paysage sonore comme méthodologie de création

La structuration du paysage sonore proposée par Schafer, distinguant une note fondamentale des autres sons, sert de fondation à des compositions musicales ou encore à des environnements génératifs. Sur la base de la composition électroacoustique, les enregistrements fixes ou mobiles des premières études du paysage sonore ont rapidement servi de matériau brut. Allant plus avant, la notion de *scène sonore* apparaît dans le contexte de reproduction ou de synthèse d'un paysage sonore. En 2006, Misra et Cook [MCW06] décrivent une scène sonore comme étant constituée d'une texture d'arrière-plan consistante avec un ou plusieurs événements sonores d'avant-plan potentiellement récurrents.

Nous abordons donc successivement la place du paysage sonore urbain dans la composition musicale, essentiellement électroacoustique, et la création d'environnements virtuels sur la base de paysages sonores urbain réels.

Paysages sonores urbains et compositions électroacoustiques

Si la musique électroacoustique n'est pas la seule à trouver inspiration dans le contexte urbain, ses techniques de réalisation impliquent souvent l'enregistrement de matériaux sonores réels, plus ou moins retravaillés par la suite en studio. Cette méthodologie est proche de celle mise en œuvre dans le design sonore. D'autre part, la diffusion des musiques électroacoustiques implique souvent l'utilisation d'un système spatialisé complexe [Mer05], incluant la notion d'espace et son appréhension par l'auditeur dans les composantes de la pièce.

La musique de paysages sonores (désignée par le terme anglais *soundscape composition*) émerge dans les années 1970, notamment sous l'impulsion du groupe de recherche *World Soundscape Project* de l'Université Simon Fraser (Canada) [Tru08]. Elle est caractérisée par la présence de sons et de contextes environnementaux reconnaissables dont l'utilisation a pour but de provoquer chez l'auditeur des associations, des souvenirs et l'imagination [Wes99]. Les techniques de composition employées sont héritées de la musique acousmatique et mettent en ap-

plication les collages sonores [Wes99], la synthèse granulaire [Tru92] et l'ensemble des traitements numériques disponibles aujourd'hui. La démarche initiale du *World Soundscape Project* était de considérer le paysage sonore comme une forme de musique, le positionnement inverse est aujourd'hui souvent utilisé pour considérer les musiques électroacoustiques comme des paysages sonores [Tru08].

Si les musiques de paysage émergent en parallèle de l'écologie sonore, elles sont également nourries de l'univers urbain. Dans un entretien daté de juillet 1999, le compositeur Pierre Henry décrit la ville comme « un réservoir de sonothèque [...], un orgue à bruits, sans silence. » Les références à l'environnement urbain sont présentes autant dans les sources utilisées que dans les titres de ses œuvres : *La ville* (1984), *Tokyo 2002* (1998) ou *Dans la rue* (2000). Ainsi, une des premières pièces affiliées à ce type de composition, « Presque rien numéro 1, le lever du jour au bord de la mer » de Luc Ferrari (1970) consiste en un assemblage de sons capturés depuis un point d'enregistrement unique (un rebord de fenêtre) pour créer « une sorte de paysage sonore ou de carte postale au sens premier du terme » [Del09]. L'ambiance perçue y est essentiellement composée de sons de moteurs, contrastant avec l'idéal promu par le titre de la pièce. Dans une approche plus complexe que celle de Luc Ferrari, les pièces de Hildegard Westerkamp mélangent les sons bruts des enregistrements et les sons traités, créant ainsi un glissement entre le monde réel et un monde imaginaire [McC99]. Il est cependant possible de débattre de la « réalité » des sons enregistrés dans la mesure où le processus d'enregistrement déforme la représentation du paysage capté, il n'en demeure pas moins que les sources restent identifiables pour l'écoutant. La pièce pour bande *A walk through the city* (Hildegard Westerkamp, 1981) illustre cette démarche avec comme point de départ des enregistrements du quartier Skid Row à Vancouver. Les composantes de la pièce sont le trafic urbain, les klaxons, les sons de freinage, de construction, de trains, des voix humaines et un poème lu par Norbert Ruebsaat. Les sources sont utilisées telles qu'enregistrées ou traitées en studio, créant un flux continu entre un paysage sonore réel et un imaginaire, tant sur le plan des matières que de l'espace qui les entoure. Hildegard Westerkamp explique que le paysage sonore de la ville moderne se forme au fil de notre changement continu de point d'attention entre ce qui est acoustiquement réel et acoustiquement imaginaire.

Plus récemment, le compositeur Thomas Köner⁵ place le thème de la ville au centre de certaines de ses œuvres audiovisuelles. *Banlieue du vide*⁶ (2003), la série *Périphériques* (2005) ou *La Barca*⁷ (2009) mettent en scène des environnements urbains intensément retravaillés sur le plan sonore et visuel.

Des outils de musique interactive mettent en pratique des principes inspirés de la musique de paysage. On peut notamment citer l'application *Scape* de Brian Eno et Peter Chilvers⁸. Cette application propose de créer de la musique en insérant des objets sur l'écran, un peu à la manière d'une carte. L'application distingue l'arrière-plan des avant-plans, chacun possédant ses propres règles de sonorisation qui permettent de créer des pièces uniques et évolutives.

La théorie du paysage sonore et le concepts circulant autour de son développement depuis les années 1970 fournissent des outils d'analyse applicables à des ensembles complexes, qu'il s'agisse de ce qu'il est possible d'entendre au quotidien ou d'enregistrements. La comparaison, par ces méthodologies, entre les paysages urbains (ou industriels) et naturels est au centre des préoccupations de l'écologie sonore et permet de produire des études définissant ce qui qualifie, perceptivement, le son urbain. Les méthodes d'analyse du paysage sonore permettent de catégoriser les sources d'un environnement en fonction de leur position relative dans l'espace et de leur évolution dans le temps, ces méthodologies sont reprises et utilisées comme des bases pour la créations de musiques électroacoustiques ou d'environnements virtuels. Ces différents concepts abordent le rôle et la signification et le sens des sons d'un environnement. Cependant, nous pensons qu'il est important d'aborder ce qui distingue un son d'un autre en fonction du timbre et quel impact peut avoir ce dernier sur la perception d'un son, d'un groupe de sons ou d'un environnement sonore complexe. Nous proposons donc dans les paragraphes suivants une brève présentation des principes de l'étude des timbres sonores et de leur application aux sons environnementaux.

5. www.koener.de

6. <http://dspace.mediaartbase.com/handle/10858/16175>

7. <http://mediaartbase.com/handle/10858/16717>

8. <http://www.generativemusic.com/scape.html>

2.1.2 Timbres sonores

À la différence des attributs auditifs de la hauteur ou du volume perçus, le timbre, n'a bénéficié de recherches que très récemment. Il se définissait principalement comme ce qui permet de distinguer deux sons présentés de manière identique, de même hauteur, de même durée subjective et de même intensité sonore [MWD⁺95].

Le timbre est un aspect perceptif du son qui se caractérise par plusieurs sous-dimensions [Gre97]. L'avènement des technologies numériques a permis d'affiner ces sous-dimensions, principalement par l'étude perceptive de timbres instrumentaux complexes.

La pluralité de ces sous-dimensions fait qu'il est impossible de mesurer un timbre sur un continuum simple (comme aigu et grave, court et long, etc.). L'un des enjeux de l'étude des timbres sonores est donc de déterminer le nombre de dimensions et de propriétés requises pour représenter les attributs perceptifs du timbre d'un son [MWD⁺95].

2.1.2.1 Espaces de timbres

De nombreuses publications tendent à reconnaître trois principales dimensions de timbre aux sons instrumentaux [Ver03] :

- La brillance, soit la distribution d'énergie spectrale corrélée au centre de gravité spectrale
- Les propriétés spectrales fines, soit la cohérence des micro-variations des composantes spectrales
- La transitoire d'attaque donnant un caractère d'explosivité ou non de l'attaque du son

Ces dimensions sont réparties dans un espace appelé *espace de timbre* [DMW94] (voir exemple Fig. 2.6) figurant une organisation mentale d'événements sonores présentant la même hauteur, la même intensité et la même durée [KMW94].

La notion d'espaces perceptifs est utilisée pour la synthèse sonore [Ver03], notamment dans l'organisation spatiale des sons pour les interfaces de contrôle gestuel.

Les principaux outils d'analyse et de classification des timbres portent sur l'étude des matériaux musicaux. Les dimensions utilisées et présentées brièvement dans les paragraphes précédents ne sont pas adaptées aux sons environnementaux car elles ne sont « représentatives que des sons continus ou impulsifs » et « ne prennent pas en compte le contexte » des sons étudiés [Ver03].

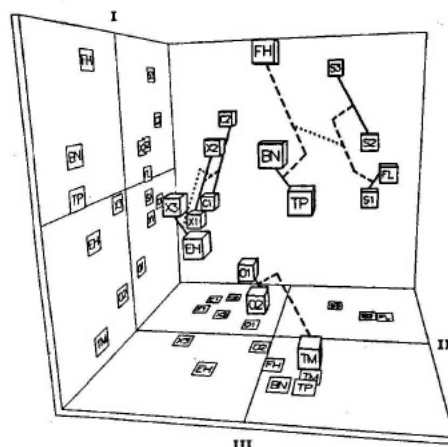


FIGURE 2.6 – Exemple de représentation d'un espace de timbre tridimensionnel extrait de l'analyse de proximité des jugements de similarité *in* [Ver03]

2.1.2.2 Timbre et sons environnementaux

De façon plus large, la recherche sur le timbre sonore doit donc être liée à la fonctionnalité du son dans notre environnement [SMM⁺05]. L'efficacité d'une information véhiculée par un son dépend de ses caractéristiques acoustiques et de sa capacité à faire reconnaître une source. L'étude des dimensions du timbre est un moyen d'isoler les dimensions perceptivement pertinentes pour la synthèse ou la création de sources sonores.

Les sons environnementaux sont produits par des actions d’agents sur des objets dans un espace donné ; cette interaction, contenue dans un intervalle temporel fini, constitue un *évènement sonore* [KB01]. Il s’agit potentiellement de tout évènement acoustique audible causé par des mouvements dans l’environnement humain ordinaire [MML⁺08]. La classification multidimensionnelle de leur timbre requiert une adaptation de la méthode d’analyse.

Pour Keller, la nature des sons instrumentaux utilisés pour les études précédemment citées exclut certaines dimensions de timbre qui seraient significatives dans l’approche des sons environnementaux. Parmi ces caractéristiques, Keller met en avant la *granularité*, l’*élasticité*, la *viscosité* et l’*état de l’objet*. La granularité semble être une dimension majeure, dans la continuité des observations de Bregman [Bre94] qui

anticipe les principaux paramètres de la synthèse granulaire appliquée aux textures : densité, durée et enveloppe des grains.

Dans une série de quatre études relatives à la description de timbres sonores [MML⁺08], Minard et al. identifient trois principales classes de sons environnementaux : les sons d'impact, les sons de moteurs et les sons ressemblant à des instruments. Chacune des ces trois catégories possède son propre espace de timbre. La *brillance* apparaît comme un descripteur commun aux trois classes, des descripteurs spécifiques sont ensuite plus significatifs pour chacune des classes séparément.

L'ensemble des propriétés mises en avant par les études sur le timbre propose des pistes de réflexion et d'analyse propres au domaine du design sonore, notamment concernant l'influence de la morphologie d'un évènement sonore sur la perception de celui-ci. Ces informations peuvent être essentielles dans la conception de sonorités visant à renseigner l'auditeur sur l'état d'un objet ou d'une scène sonore. Elles sont complémentaires des analyses du paysage sonore qui fournissent des méthodologies de structuration temporelle et spatiale des évènements sonores entre eux dans le cas de la composition d'une scène environnementale.

Dans la section suivante, nous proposons de considérer la ville comme un paysage sonore et d'en appréhender les principales sources le composant ainsi que les phénomènes acoustiques et perceptifs liant les sources à l'environnement pour l'auditeur. Le principal objectif de cette section est de mettre en avant le concept d'arrière-plan sonore urbain, inhérent à la ville sonore, et d'en extraire les principales caractéristiques acoustiques et perceptives.

2.2 La ville sonore

Dans cette section nous abordons la ville sonore au travers d'analyses acoustiques et perceptives. Le paysage sonore urbain est la résultante de l'activité humaine dans l'espace public [BDD06]. Il est aujourd'hui, le plus souvent, dominé par le son du trafic routier qui masque les autres composantes de la vie urbaine. Malgré des différences culturelles observables d'un pays ou d'un continent à un autre, le trafic exerce une « emprise » sur le paysage sonore urbain et en sature le contenu [Tor05]. L'intégration des concepts de paysage sonore à la planification urbaine a pour objectif de rompre avec la dominante des sons du trafic, au profit d'une diversité sonore et d'une perspective acoustique accrues [BDD06].

Henry Torgue [Tor05] définit le son urbain comme le mélange de trois phénomènes : des *fonds*, des *séquences* et des *signaux-événements* :

- Les *fonds* sont des « états sonores stables » qui se modifient lentement tout en caractérisant de grands axes ou de grands espaces. En milieu urbain, les principales composantes en sont les transports et les activités industrielles qui forment la base de ce drone. Le drone est composé majoritairement de fréquences graves, il est la « rumeur sourde et intense » de la ville.
- Les *séquences* sont des marqueurs spatiotemporels liés à des activités ou à des circulations localisés qui se superposent en plans sonores.
- Les *signaux-événements* constituent les sources perçues comme ponctuelles (cloche, avertisseur, etc.) et qui attirent l'attention de l'auditeur.

Torgue insiste sur la relativité de ces catégories selon la position de l'écouter, tant dans l'espace (distance par rapport aux sources) que dans son rapport à l'environnement (familiarité ou passage). La ville sonore est donc une interaction entre des sources, des espaces de diffusion et des perceptions « liées à des critères sensibles ». Les sources, décrites précédemment par le biais de leur regroupement en trois catégories, se caractérisent par leur nombre très élevé en milieu urbain. Les espaces de diffusion sont les espaces construits et naturels dans lesquels se propagent les sources. Les perceptions recouvrent les significations, les représentations et les interprétations des signaux perçus dans les espaces de diffusion.

Les interactions entre ces trois dimensions sont à la base de la complexité de la diversité de la ville sonore. Pour Torgue, ces interactions se retrouvent à toutes les échelles de perception de la ville, « de l'intimité d'un logement à la géographie d'ensemble d'une agglomération ».

Le travail de thèse présenté ici cherche principalement à appréhender la notion d'arrière-plan, dans l'environnement urbain, telle que définie dans la théorie du paysage sonore et par Henry Torgue. Comment sonne la ville ? Quel est la place de l'arrière-plan dans le paysage sonore urbain ? De quoi ce bruit de fond est-il composé ? Comment est-il perçu ? Les paragraphes suivants tentent de répondre à ces questions en abordant tout d'abord les mécanismes de la propagation sonore en milieu urbain puis des études perceptives sur la ville sonore.

2.2.1 Analyse acoustique

Afin de cerner l'objet d'étude qu'est l'arrière-plan sonore, il paraît important de présenter brièvement les mécanismes acoustiques propres à la ville et en quoi ils contribuent à la création d'un bruit de fond spécifique.

Une méthode pour analyser le paysage sonore urbain est d'en confronter les caractéristiques directement observables avec celles d'autres paysages. Dans une étude de 2006, Botteldooren et al. [BDD06] distinguent ainsi les paysages sonores urbains, ruraux et naturels. Ils montrent (Fig. 2.7) les différences entre ces catégories en terme de variation d'intensité du volume (soit la dynamique sonore globale) et de tonalité (soit une certaine musicalité globale). Les caractéristiques qui en ressortent concernant le son urbain sont une dynamique et une tonalité bien plus constantes que dans les paysages naturels et ruraux. Ces observations rejoignent celles présentant le paysage sonore urbain comme *lo-fi* [Tru08] (voir 2.1.1), c'est-à-dire comme un paysage souvent uniforme, dont le volume sonore général et le contenu harmonique perçus varient peu ou lentement au fil de la journée.

De nombreuses recherches portent sur la simulation des mécanismes de propagation du son en milieu urbain [Kan02]. Ces recherches montrent que les ondes sonores tendent à se réfléchir de nombreuses fois sur l'environnement avant d'atteindre l'auditeur. Chaque réflexion modifie le contenu fréquentiel de l'onde sonore de départ ainsi que la perception de sa localisation dans l'espace. L'arrière-plan sonore n'est pas la résul-

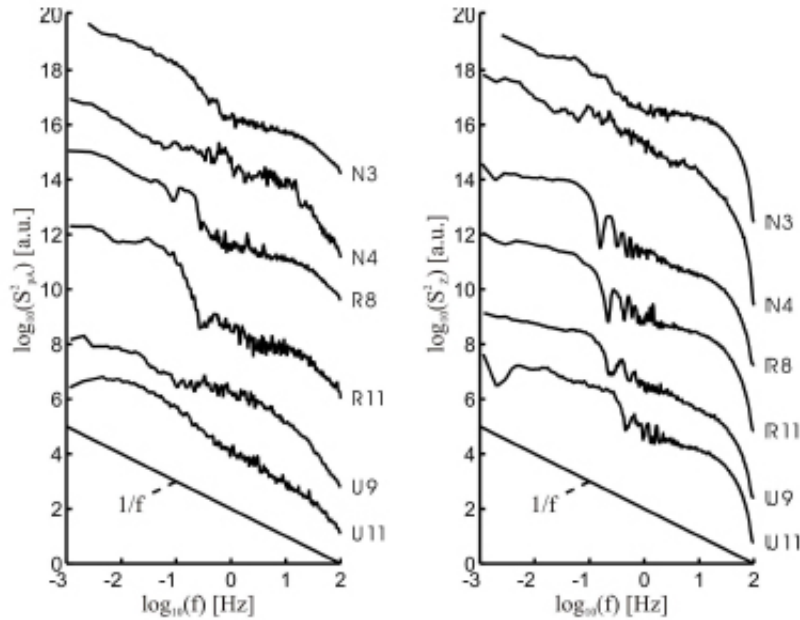


FIGURE 2.7 – Fluctuations de l’amplitude (gauche) et de la tonalité (droite) pour les paysages sonores naturels (N3 et N4), ruraux (R8 et R11) et urbains (U9 et U11), *in* [BDD06]

tante d’une source sonore précisément identifiable mais la conséquence d’une accumulation de réflexions acoustiques sur l’environnement. Il se compose essentiellement de la combinaison de deux éléments : les sons de trafic réfléchis par les façades du bâti.

Nous n’aborderons pas ici les différents mécanismes analysés ni leurs méthodes de simulation mais le rôle de ces mécanismes dans la création du bruit de fond urbain ainsi que des analyses dédiées à son contenu spectral. Les paragraphes suivants traitent donc de la notion de champs sonores et de leur formation dans l’environnement urbain puis des études portant sur le contenu fréquentiel de l’arrière-plan sonore urbain.

2.2.1.1 Champs sonores

Le son perçu dans un environnement est le plus souvent un mélange entre le *champ direct* (soit une ligne droite entre la source et le point d’écoute) et le *champ indirect* (ou *diffus*, composé des réflexions de la

source sur l'environnement et parvenant au point d'écoute, Fig. 2.8). La balance entre les champs diffus et direct est liée à la position du point d'écoute par rapport à la source ainsi qu'aux éventuels obstacles se trouvant entre eux.

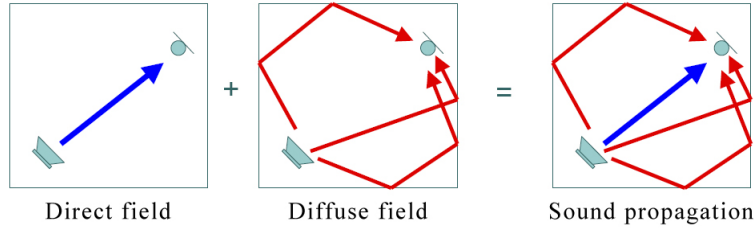


FIGURE 2.8 – Mélange des champs acoustiques direct et indirect issus d'une source sonore et formant l'ensemble perçu au point d'écoute

Les mécanismes acoustiques de la ville sont en grande partie dus à la disposition rectiligne des façades sur lesquelles viennent se réfléchir et se diffuser les ondes sonores. On parle de l'*effet canyon* [Kan00] pour décrire ce phénomène (Fig. 2.9). La manière dont les ondes sonores se propagent à l'intérieur du canyon est fortement liée au type des façades. Dans les grandes villes, celles-ci présentent majoritairement des irrégularités, typiques d'immeubles anciens, qui tendent à disperser aléatoirement le son [IO05].

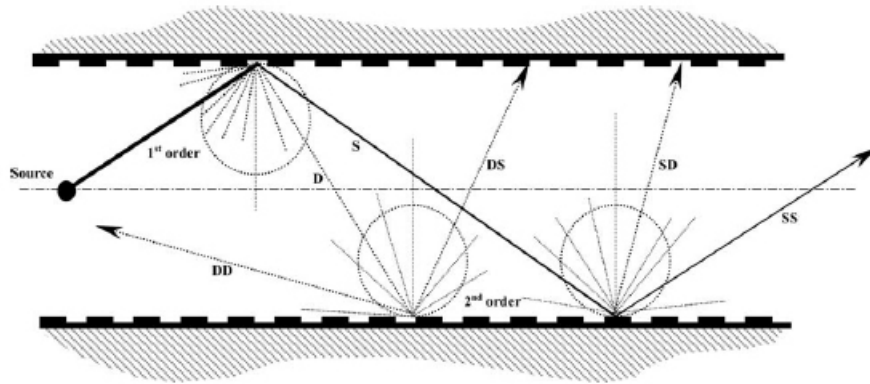


FIGURE 2.9 – Représentation de l'effet canyon ainsi que de la dispersion acoustique d'une source sur les façades, *in* [IO05]

Le fond sonore urbain est un champs diffus causé par la multitude des réflexions acoustiques sur l'environnement [PHS99]. Le nombre im-

portant de sources sonores présentes crée une densité suffisante de réflexions pour générer un bruit de fond permanent et faiblement localisable [WMG65].

De manière générale, les fréquences graves tendent à se propager plus facilement alors que les fréquences aiguës sont plus diffusées et dispersées [Heu09]. Ainsi, le spectre du bruit de fond urbain perd rapidement en hautes fréquences tandis que les fréquences graves tendent à persister.

2.2.1.2 Spectre de l'arrière-plan urbain

Comme nous l'avons vu, la principale source sonore en milieu urbain est le trafic automobile. Les moteurs et les sons de roulement en sont les principales composantes. Dans une étude de 2010, Can et al. ont effectué des relevés acoustiques de la circulation aux abords du Cours Lafayette (Lyon) [CLLB10]. Les relevés, effectués simultanément depuis plusieurs points d'enregistrement, mettent en lumière certaines propriétés du spectre sonore urbain.

Les mesures et analyses spectrales des ambiances urbaines montrent dans leur ensemble une baisse d'intensité du grave vers l'aiguë, autour de 8000Hz. Cette baisse d'intensité n'est pas régulière, elle présente une bosse entre 1000Hz et 2000Hz, ce qui correspond au pic observé dans les analyses de sons du roulement des pneus sur la chaussée [San03]. Les analyses fréquentielles de cette étude montrent également une grande variabilité d'intensité selon le point d'écoute. Un déplacement reculant le point d'écoute de quelques mètres réduit significativement l'intensité perçue. Comme exposé dans le point précédent, les fréquences aiguës perdent davantage en intensité que les fréquences graves.

Une étude de Shaw et Olson [SO72] portant spécifiquement sur la composition fréquentielle de l'arrière-plan sonore urbain propose un schéma de forme très semblable avec davantage de filtrage des fréquences aiguës (Fig. 2.10).

Les propriétés sonores du trafic interagissent avec les propriétés acoustiques de l'environnement urbain. En se propageant dans les rues, les strates réverbérées s'accumulent, créant un bourdon dense dont la composante spectrale varie selon l'intensité du trafic et la position du point d'écoute. Cette accumulation devient un objet sonore, une texture à part entière, comme une empreinte de la ville.

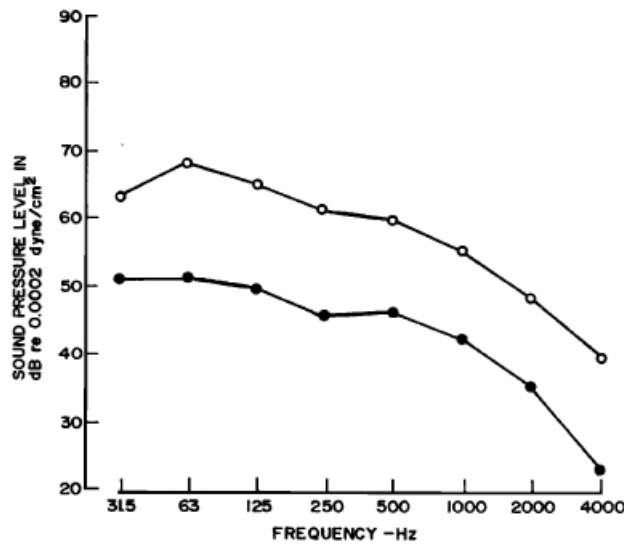


FIGURE 2.10 – Comparaison de la répartition des fréquences entre les arrière-plans sonores urbains enregistrés (points noirs) et simulés (points blancs), présentée dans [Lyo74] selon le modèle proposé dans [SO72]

La quantité importante de réflexions le long des façades réduit également la lisibilité spatiale de la scène, ainsi, les objets lointains qui ne sont perçus que par le biais du champ indirect deviennent difficilement localisables. L'ensemble des réflexions de toutes les sources crée une masse sonore compacte et ubiquitaire.

Nous avons vu, dans cette sous-section, que l'arrière-plan sonore urbain trouve son origine dans les mécanismes de diffusions acoustiques de la ville, l'accumulation de ces diffusions créant un drone persistant. La coloration fréquentielle de ce drone est due à l'interaction entre les sources et la forme de l'environnement. Cependant, des constantes semblent exister quant au plus fort filtrage des fréquences aiguës par rapport aux fréquences graves. Le timbre du fond sonore urbain apparaît ainsi peu brillant et faible en transitoires.

Dans la sous-section suivante, nous abordons la ville sonore analysée perceptivement. Les études perceptives de la ville nous permettent de présenter les sources sonores peuplant le paysage sonore au travers d'approches sémantiques ainsi que la présence de l'arrière-plan sonore.

2.2.2 Analyse perceptive

L'analyse des mécanismes acoustiques propres à l'environnement urbain est un support objectif complémentaire de la théorie du paysage sonore. Cependant, l'approche analytique ne peut rendre compte de la manière dont l'environnement urbain est perçu par ses habitants. La perception de l'environnement urbain est en effet liée à d'autres facteurs, subjectifs, liés à l'expérience faite de la ville et à l'imaginaire sonore qui s'y rattache [Che96].

Les études relatives à la perception de l'environnement sonore urbain ont le plus souvent pour objectif la mesure de la *qualité sonore*. La notion de qualité sonore [MCD98] permet d'appréhender différentes facettes, qualitatives ou sémantiques, de la perception. Elle est souvent le point de départ d'une réflexion plus globale sur la place de l'écologie sonore dans les plans d'urbanisme. La notion de qualité sonore peut faire référence tant à l'interprétation des niveaux de bruit objectivement mesurés [SMW99] qu'à l'agréabilité perçue subjectivement [DPL⁺07]. Ces approches multidimensionnelles qualitatives de l'environnement urbain amènent des classifications sémantiques des scènes qui le composent. Les classifications sont des outils d'étude de la ville sonore, ils sont utiles au regard de sa complexité perceptive : elles permettent d'identifier ce qui est entendu et la signification de ce qui est entendu de la ville.

Après avoir souligné la complexité perceptive de la ville sonore, nous tenterons de répondre aux questions suivantes : Quel est le principe des classifications sonores ? Comment sont-elles appliquées à la ville ? Que disent-elles des fonds sonores urbains ?

2.2.2.1 Complexité perceptive de la ville sonore

La complexité de la structuration de la ville ainsi que la densité des sources qui l'habitent [Vog99] renvoient à des problématiques perceptives. Deux champs de recherche de psychologie cognitive se rapprochent des enjeux de l'étude de la perception sonore en milieu urbain. La première est l'*analyse de scènes auditives* [Bre94], qui étudie la perception du nombre, des caractéristiques et du positionnement des sources sonores dans un espace. La seconde est l'effet *cocktail party* [Che53], soit la capacité d'un individu à porter son attention sur un interlocuteur parmi la cacophonie des conversations et le bruit de fond [Aro92]. Ces différentes approches tentent de comprendre les méca-

nismes et les limites de la perception sonore dans les environnements complexes. La thèse d'Antoine Gonot [Gon08] propose un état de l'art sur l'influence de l'espace et de la spatialisation dans la perception sonore.

Les notions d'acoustique et de paysage sonore que nous avons vues peuvent être complétées par celle d'*ambiance sonore* développée au laboratoire CRESSON [ATC04]. Le concept d'ambiance se différencie des approches acoustiques (dominées par la technique et la notion de bruit) et paysagères (qui entendent l'environnement dans un sens quasi musical) en privilégiant une approche transversale liée à l'*espace* («acoustique appliquée, architecture, urbanisme»), la *perception* («psychologie et sociologie du quotidien, imaginaire») et la *production* sonore («conduites individuelles et collectives, nouvelles techniques, communication»). La qualité sonore des espaces habités «est en effet liée à l'interaction de ces trois champs». L'approche des ambiances vise à comprendre la ville sonore en interrogeant les «relations actives à l'environnement, ou la constitution du milieu d'action et d'interaction» [Che96] en faisant de la perception un élément central [Thi02]. Le phénomène sonore agit comme un révélateur de la ville en «éveillant» l'acoustique du bâti par les mécanismes de propagation et en témoignant des actions qui sont à l'origine de ces phénomènes. L'approche théorique de l'ambiance urbaine propose également une définition du phénomène de l'arrière-plan sonore urbain. L'arrière-plan sonore urbain se superpose aux sons identifiables individuellement. Il est ce qu'il reste dans le silence paradoxal de la ville, bruit de fond insensible tant il est habituel [Che96]. Il a deux effets principaux sur la perception de l'environnement : il limite la perspective acoustique en agissant comme un masque sur les sources lointaines et, parallèlement, uniformise la perception sonore des territoire. Dans [Che96], Grégoire Chelkoff reprend les observations du futuriste Luigi Russolo [Rus03] définissant le bruit de fond urbain comme :

(...) une basse continue et soutenue qui agit comme pédale pour tous les autres bruits. Il n'est pas facile de relever les caractéristiques tonales de ce bruit : pourtant plusieurs fois, j'ai réussi à le caractériser comme un accord parfait assez clair ; d'autres fois au contraire, comme une quinte seulement (...)

La compréhension et la structuration des composantes sonores de la

ville est une étape primordiale à sa recomposition virtuelle car cela permet de déterminer une stratégie de sonorisation : quoi faire entendre, sous quelle forme et selon quelles contraintes temporelles, spatiales ou événementielles.

2.2.2.2 Objets sonores urbains

Dans les sections précédentes, nous avons vu comment la théorie du paysage propose de segmenter une scène en différents plans. Les recherches sur les classifications perceptives des sons, que nous allons aborder dans les paragraphes suivants, offrent des granularités plus fines de classification. Ces classifications utilisent différentes approches (morphologiques, sémantiques,...), elles permettent de dresser une liste des sources sonores présentes dans l'environnement et d'appréhender la manière dont ces sources sont perçues et interprétées. Certaines de ces classifications sont réalisées spécialement pour l'analyse de l'environnement urbain, leur étude nous paraît pertinente en vue du développement d'outils de design sonore dédiés aux villes virtuelles.

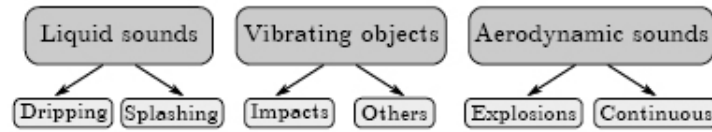
La principale méthode de classification des sons consiste à dresser des taxonomies soit par l'intermédiaire de questionnaires, d'observations subjectives ou d'analyses d'enregistrements.

Dans la continuité de la théorie du paysage sonore, Schafer propose une classification référentielle des sources sonores [MS10]. Cette classification est basée sur le contexte des sources davantage que sur leur contenu ou leur description physique et propose les catégories suivantes :

- *Natural sounds* (sons naturels)
- *Human sounds* (sons humains)
- *Sounds and society* (sons des environnements sociaux)
- *Mechanical sounds* (sons mécaniques)
- *Quiet and Silence* (calme et silence)
- *Sounds as indicators* (indicateurs sonores)

Si cette taxonomie propose de catégoriser les sons selon la nature de leur source, celle proposée par Gaver [Gav93] met en avant des aspects plus morphologiques en distinguant les sons liquides, aérodynamiques et ceux provenant d'objets en vibration (Fig. 2.11).

La recherche de qualité sonore dans l'environnement urbain est souvent conduite dans une approche quantitative (relevés en décibels, analyses spectrales). Cependant, l'environnement sonore urbain dépend en

FIGURE 2.11 – Simplification de la taxonomie de Gaver, *in* [VAKMP10]

grande partie de pratiques sociales bruyantes [Che96] ce qui limite les observations quantitatives à des mesures ne reflétant pas la nature des phénomènes complexes qui en sont la cause. Le quantitatif montre donc ses limites et se complète de nouveaux concepts (ambiances, paysages) qui introduisent des facteurs qualitatifs et perceptifs. Il devient alors nécessaire de concevoir des études qui portent non plus uniquement sur la notion de bruit, analysée par des experts, mais qui cherchent à décrypter comment l'environnement urbain est perçu, compris et catégorisé par ses habitants. Ces études ont recours à différentes méthodologies telles que l'analyse perceptive d'enregistrements, des questionnaires adressés aux habitants. Un certain nombre d'approches visant à classer et catégoriser spécifiquement les sons urbains se développent sur la base des méthodes perceptives [GKW07] de catégorisation des sons environnementaux [MMHS10], ces méthodes étant parfois déduites de celles utilisées pour l'analyse musicale [MMS⁺10]. D'autres méthodologies basées sur l'observation directe sont notables. Ainsi, les *soundwalks* sont par exemple utilisées comme des outils d'évaluation à destination des professionnels de la planification urbaine [ABD⁺08]. Il s'agit de faire dialoguer les enjeux soulevés par les études qualitatives avec les transformations futures du tissu urbain [VS06].

L'approche environnementale du bruit, dans la planification urbaine, ne consiste pas uniquement à en réduire l'intensité (c'est-à-dire de le réduire à une nuisance), il s'agit également de comprendre les informations utiles fournies par le bruit. Avec pour objectif de caractériser les critères perceptifs liés à la qualité sonore, Susini et al. [SMWM98] comparent deux études qui proposent de catégoriser, dans un premier cas, différents sons de moteur et, dans un second cas, des sons urbains appartenant à différentes « familles » (rue, autoroute, aéroport, entre autres). La comparaison entre les deux expériences montre que les mécanismes perceptifs mis en œuvres par les écoutants

pour catégoriser les sons ne sont pas identiques. La disparité du corpus d'échantillons sonores de la deuxième expérience force l'identification et la relation directe du son à l'objet, ce qui empêche le recours à des mécanismes de perception selon des traits auditifs abstraits. Dans la deuxième expérience l'attention de l'écouter se porte sur l'identification des différentes familles sonores du corpus alors que, dans la première expérience, l'appartenance des sons à la même famille (sons de moteurs) permet à l'écouter de classer le corpus selon des critères plus abstraits. Les auteurs expliquent cette différence de la façon suivante :

La comparaison des résultats de ces deux types d'études permet de mesurer en partie l'influence de la nature des sons sur les structures perceptives qui sous-tendent la perception auditive. En effet, il semble que si les stimuli sonores sont assez similaires, donc difficilement séparables en classes distinctes, leur représentation mentale est liée à différents attributs auditifs continus que l'on peut corrélérer à des paramètres physiques des signaux. Par contre, si le corpus de stimuli est de nature extrêmement hétérogène, le facteur cognitif lié à l'identification des sources impose une structure perceptive catégorielle et empêche inévitablement le jugement à partir de critères perceptifs plus fins.

Si certaines études tendent vers une classification automatique des sons environnementaux [CNK09], les auteurs affirment ici que le meilleur instrument de mesure de la notion de qualité sonore est l'auditeur humain, avec le recours de la psychoacoustique et des méthodologies de psychologie expérimentale qui ne peuvent ici ignorer les « facteurs cognitifs liés aux mesures subjectives. »

La possibilité d'identifier les sources constitue un critère essentiel dans la classification perceptive. En effet, les attributs servant à décrire des phénomènes sonores urbains [Gua07] varient selon l'identification de leur source ou de leur cause interprétée. De même, il est possible de noter que l'intérêt porté à un paysage sonore urbain peut en faire varier l'interprétation et la catégorisation du contenu. En 2005, Raimbault et Dubois [RD05] montrent les différences (Fig. 2.12) entre les descriptions d'un même paysage sonore urbain selon qu'il soit entendu par ses habitants ou par des planificateurs urbains.

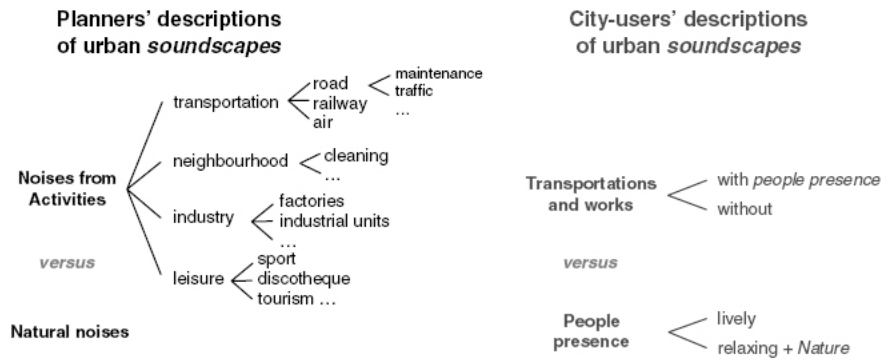


FIGURE 2.12 – Comparaison entre les descriptions verbales de paysage sonore urbain fournies par ses habitants et par des planificateurs urbains in [RD05]

Ces deux classifications divisent, selon le point d'intérêt de l'écoute, le paysage sonore entre *bruits naturels* et *bruits d'activités* (transports, usines...) ou selon la présence du facteur humain. Les auteurs de l'article expliquent que les catégories mises en avant par les professionnels sont directement en lien avec leur profession et le concept de nuisance sonore, la perception et la catégorisation du contenu d'un paysage sonore urbain est donc aussi soumise à des facteurs subjectifs.

Les études que nous avons brièvement présentées font ressortir la part active de l'auditeur dans la perception de la ville sonore. L'identification et la classification des sources sonores semble liées à un contexte et à une intention dans l'interprétation de la ville sonore. Ces éléments peuvent apporter des pistes de réflexion dans la conception sonore d'une ville virtuelle dans la mesure où ce qui est donné à entendre est susceptible d'influencer l'auditeur sur sa perception et son interprétation de l'environnement virtuel. Notons qu'il n'existe pas encore de consensus concernant la classification sémantique des objets sonores urbains, même si des études récentes tendent à aller dans ce sens [SJB14]. Dans les paragraphes suivants, nous présentons des études de catégorisation faisant ressortir l'arrière-plan sonore comme une entité du paysage urbain.

2.2.2.3 Présence de l'arrière-plan sonore

Maffiolo et al. [MCD98] procèdent à une étude perceptive afin de catégoriser des scènes urbaines (soit des « fragments de l'environnement sonore porteur de sens et d'identité » d'une durée de quinze à vingt secondes chacun). Les méthodologies mises au point pour l'enregistrement et l'écoute des scènes permettent de distinguer, après catégorisation libre, deux principaux types de scènes sonores : les scènes *événementielles* contenant des événements imputables à des sources identifiables et les scènes *amorphes* constituées de bruit de fond sans événement spécifique. Cette distinction met en avant la différence de traitement perceptif d'une scène urbaine selon son contenu, les scènes événementielles étant verbalement décrites plus facilement par leurs événements identifiables tandis que les scènes amorphes sont verbalement décrites selon leurs propriétés physiques et semblent dénuées de sens.

Dans une étude de 2006 portant sur la perception du son de la ville par ses habitants et basée sur des questionnaires [Gua06], une distinction similaire apparaît entre les événements sonores (pouvant être attribués à une source ou un agent identifié) et le bruit d'arrière-plan urbain (un son d'ensemble dont aucune source particulière n'émerge). Les descriptions des habitants relatives au bruit de fond se basent également principalement sur des descriptions physiques du signal (structure temporelle et timbre), ce qui suggère une conceptualisation et un traitement cognitif plus abstraits du bruit d'arrière-plan en tant qu'objet sonore. Du point de vue de l'appréciation, l'arrière-plan est mieux accepté s'il est temporellement continu (par opposition à un bruit à la dynamique irrégulière), ceci est renforcé par la sensations de sources d'origine humaine [DGR06], par opposition à la sensation d'un bruit de fond d'origine uniquement mécanique.

Les études que nous avons passées en revue dans cette section consacrée à la ville sonore font ressortir la complexité acoustique du milieu urbain, en raison du nombre de sources le peuplant et de sa morphologie. Les diffusions acoustiques importantes sont propices à la création d'un fond sonore continu. Les différentes sources sonores peuplant la ville sont perçues et catégorisées selon des critères parfois subjectifs et relatifs au rapport de l'écoutant à la ville. Perceptivement, les arrière-plans sonores sont décrits de façon plus abstraite que les sources indi-

viduellement identifiables, en plus de leur stabilité dans le temps, ils sont caractérisés par une absence d'événement associé. Des études plus récentes [SJB14] attestent également de la différence significative entre premier et arrière plans dans la perception de l'environnement sonore urbain.

Nous pouvons alors nous demander comment ces facteurs acoustiques et perceptifs sont pris en compte lorsqu'il s'agit de reproduire le son de l'environnement urbain. Le cinéma, les simulations ou les jeux vidéo mettent régulièrement en scène des paysages urbains. Quelles sont alors les stratégies mises en œuvre pour le reconstituer ? La section suivante tente de répondre à cette question en abordant successivement les représentations audiovisuelles linéaires et interactives de la ville.

2.3 Virtualisation et sonorisation de la ville

Dans les sections précédentes, nous avons vu que l'arrière-plan est une composante de l'environnement sonore. Dans cette section, nous interrogeons la place de l'arrière-plan sonore dans les représentations audiovisuelles de l'environnement urbain, sur la manière dont il est représenté ainsi que sur les significations qu'il recouvre dans le processus de l'écriture sonore de la ville audiovisuelle.

Cette section porte dans un premier temps sur la ville sonore représentée au cinéma. Le cinéma se développe en parallèle du cadre urbain moderne et propose la première représentation audiovisuelle de la ville, il définit une partie des règles de sonorisation utilisées par la suite.

Les villes sonores au cinéma constituent donc un préambule à l'étude de la sonorisation des villes virtuelles interactives. Ces dernières sont au centre de notre travail de thèse, l'étude de leur mécanisme de sonorisation nous apparaît essentielle dans la perspective de concevoir un outil de synthèse interactif temps-réel dédié aux arrière-plans sonores urbains.

2.3.1 Villes sonores au cinéma

Nous traitons ici brièvement des représentation audiovisuelles linéaires de la ville. Ces représentations vont de la perception du réel (comment reconstituer la réalité de la ville sonore ?) à l'abstraction (quelles représentations symboliques de la ville sonore ?). Toutes posent en définitive également la question du hors-champ et donc de la manière dont l'espace entendu se superpose à l'espace vu, selon quelles nuances, et comment cette superposition influence la signification de la ville pour le spectateur. Comment les villes sont-elles représentées au cinéma ? Quelle est la fonction des ambiances urbaines au cinéma ? Le paysage sonore urbain au cinéma tend-il vers une uniformisation ou bien cherche-t-on au contraire à distinguer les villes les unes des autres ?

Pour Michel Chion [Chi94], la principale fonction du son au cinéma est d'unifier ou de lier le flux d'images selon trois perspectives. Tout d'abord, sur un plan temporel, le son crée une continuité entre les plans visuels. Ensuite, le son crée des atmosphères et des espaces qui constituent un support pour le visuel. Enfin, la musique non-diégétique,

du fait de son indépendance temporelle et spatiale, est à même de projeter les images dans un ensemble homogénéisé.

Dans la construction cinématographique, le son, en raison de son lien avec l'espace environnant, exprime trois caractéristiques principales [BRFH⁺11] : le contexte spatial, la synchronicité et les rythmes. Ces trois caractéristiques sont librement influençables selon le discours de la réalisation. Les éléments qui composent ces principales caractéristiques, y compris les ambiances et arrière-plans, peuvent être utilisés comme des leitmotivs. L'acte de sonorisation permet aussi de créer un effet de zoom sur les sons du quotidien : des sons proches du silence sont souvent magnifiés à l'écran, bien plus intenses que dans les conditions réelles.

Le langage sonore cinématographique s'articule entre les sons dits *diégétiques* et *non-diégétiques*. Parmi les sons diégétiques, Chion [Chi94] distingue les sons visibles et invisibles. Les premiers sont ceux dont la source apparaît à l'écran, les seconds sont relatifs à ce qui est visible mais dont la source est hors-champ (Chion parle également ici de sons acousmatiques, c'est-à-dire destinés à être entendus sans que leur cause ne soit visible). Les sons non-diégétiques sont ceux dont la source est absente de l'image et n'a pas de lien direct avec le déroulement de l'histoire ou de l'action.

Ces distinctions sont un outil fonctionnel pour analyser la *topologie sonore* d'un film. Chion propose d'y ajouter certaines catégories afin d'affiner l'analyse lorsque cela est nécessaire : le son *ambient*, le son *interne* et le son *on-the-air*. Le son *ambient* trouve un écho plus particulier au sujet traité dans cette thèse. Chion lui donne également l'appellation de *son territoire* car il sert à identifier une zone, un lieu, par sa présence continue. Le son *ambient* rend également compte de l'espace d'un environnement ainsi que la manière dont cet espace est habité du point de vue sonore. Chion inclut le son *ambient* dans le son *hors-champ passif* (par opposition au son *hors-champ actif*), ce dernier créant une atmosphère vouée à envelopper et stabiliser l'image, sans distraire le regard ou inciter à chercher la source. Le son *hors-champ passif* est un élément sans dynamique et stable sur lequel repose l'édition de l'action. Ce support (Chion cite ici en exemple le mélange général de l'ensemble des sons d'une ville) permet à la caméra de se déplacer plus librement dans l'espace, d'inclure, par exemple, davantage de gros plans, tout en conservant la cohérence du lieu et du temps. Cette définition rappelle fortement celle de la notion d'arrière-plan dans la

théorie du paysage sonore, par la constance dynamique, la faculté de caractériser un espace tout en constituant un support pour les autres entités sonores de l'environnement.

Les villes cinématographiques créent des représentations collectives, génériques et ancrées dans l'imaginaire par l'utilisation de *soundmarks* permettant d'identifier une ville par un élément sonore (par exemple les sirènes de New York). Le son a un rôle central dans la représentation de la ville. De manière générale, lorsqu'une bande-son est construite comme un ensemble stable, elle peut être un vecteur auquel l'auditeur se raccroche si les plans changent et se coupent rapidement, déplaçant ainsi le spectateur dans différents endroits de l'espace du film [Ros12]. La bande-son inclut donc différents points d'écoute de la ville tout en maintenant son unité et son identité sonore.

Cette section traitera tout d'abord du cas des films de type *City Symphonies* puis plus spécifiquement du rapport entre la ville filmée et ses ambiances urbaines.

2.3.1.1 *City symphonies*

Si la ville moderne se développe en parallèle du cinéma [MMG09], elle n'en reste pas moins en second plan à l'écran, constituant le plus souvent un cadre de narration. Lorsque le cas inverse se présente, l'argument et les personnages sont mis en retrait au profit d'une chronique de l'urbain. C'est le cas des *city symphonies*, une série de films des années 1920 et 1930 produits à la charnière des cinémas « muets » et « parlants ». Parmi les films emblématiques de cette vague, citons *Berlin, symphony of a great city* (1927, Fig. 2.13), *Man with a movie camera* (1929), *São Paulo, a Metropolitan Symphony* (1929) ou bien encore *Rien que les heures* (1926) [MdC12].

La projection de ces films n'était pas « muette » mais accompagnée d'une bande-son étudiée pour l'image. Il pouvait s'agir d'une bande-son pré-enregistrée et synchronisée (par exemple *Berlin, symphony of a great city*) « accentuant la valeur expressive des images, comme il est encore habituel dans les bandes sonores » [MMG09]. Le langage sonore inclut des parties orchestrales, des tempos synchronisés sur les machines montrées à l'image, des recherches de timbres instrumentaux pour se rapprocher de ceux des machines. L'utilisation de crescendos accentue les effets visuels et la dynamique des actions présentes



FIGURE 2.13 – Image extraite du film *Berlin, symphony of a great city* (1927)

à l'écran. Il s'agit de représenter, au travers de la musique, la sensation des actions dont les sons correspondants n'étaient pas enregistrés.

Les techniques de production avançant, des enregistrements de sons quotidiens se sont mêlés à la musique, accompagnant, parfois rythmiquement, le visuel (*Melody of the world*, 1929). Cette mixité est particulièrement présente dans *Entuziazm* réalisé par Dziga Vertov en 1931 : le film introduit un chef d'orchestre dirigeant sa propre bande-son, cette dernière incluant la musique mais également des sons du quotidien urbain (cloches, usines, foules, machines ou encore trains). Les principes mis en œuvre par Vertov dans la production de ses films trouvent leur origine dans le mouvement Futuriste [Nor84], il s'agit de mettre en avant le pouvoir de la technologie (la supériorité de l'œil mécanique de la caméra sur l'œil humain) et du mouvement. Pour l'incorporation de ces enregistrements dans la bande-son de *Entuziazm*, Dziga Vertov utilise le premier enregistreur portable de brevet russe dont la taille réduite permet d'effectuer des enregistrements en extérieur. Cette recherche sonore est un prolongement du « cinéma-vérité » recherché au travers d'images authentiques s'opposant aux décors des studios [MMG09]. Vertov articule un langage sonore symbolique et complexe mélangeant la machine et l'humain ; ce langage est composé de nombreux effets [Fis77] parmi lesquels des surimpressions, des inversions, des éditions franches, des contrastes de tonalité ou des dissonances spatiales.

Dans les *city symphonies*, une attention toute particulière est portée sur l'aspect rythmique, quasi métronomique du son urbain. Ces découpages rythmiques inspirés des machines se mélangent à des composantes bruitées et des masses sonores inspirées des foules et de la rumeur du trafic dont on ne peut distinguer les composantes individuelles.

Les méthodes proposées par les *city symphonies* restent aujourd'hui d'actualité pour la sonorisation des villes filmées. Un mélange de sons produits en studio et d'enregistrements réalisés sur le lieu de tournage ainsi que des effets plus métaphoriques constituent le langage de base de la sonorisation de l'urbain filmé. Les ambiances urbaines permettent de donner une coloration tonale et d'humeur, en complément ou substitution de la musique, tout en apportant une consolidation de l'espace.

2.3.1.2 Villes et ambiances

Les ambiances sonores, urbaines ou non, sont des constructions plus ou moins artificielles, oscillant entre le son direct, enregistré en parallèle de l'image, et la reconstitution ou création complète en studio à partir d'enregistrements tiers [Chi94]. Cependant, cette construction est nécessairement contrôlée afin de conserver l'intelligibilité des premiers plans, c'est-à-dire le plus souvent des dialogues. Les ambiances urbaines sont donc un mélange de rumeurs (trafic), de masses sonores (foules) et d'événements (claquements de portières, klaxons, sirènes). Ces différents éléments peuvent apporter plus ou moins d'informations sur la localisation, l'identité ou l'anonymat du cadre urbain, selon les intentions de réalisation.

Michel Chion prend l'exemple du film *The aviator's wife* (1981) dans lequel certaines scènes tournées dans Paris utilisent des ambiances tant épurées qu'elles en deviennent neutres et anonymes. Ces ambiances sont pourtant enregistrées en direct mais avec le souci de ne capter aucun son anecdotique. Pour arriver à un tel résultat, le réalisateur a délibérément tourné à des heures et en des endroits qui présentaient la moindre exposition sonore, il a également rejeté les prises contenant des intrusions. Cette démarche est à l'inverse de celle choisie par Jacques Tati mettant en scène un bruit urbain ubiquitaire envahissant la narration [Laf07]. Pour Michel Chion, il s'agit donc d'un

processus de « choix et élimination » qui confère à la bande-son de *The aviator's wife* un caractère non moins reconstruit que la post-synchronisation. Michel Chion soulève ici le paradoxe suivant : l'élimination de tout son qui pourrait perturber l'attention du spectateur élimine également une partie de la vie entourant l'action. Un certain nombre de sons anecdotiques sont attendus par le spectateur, ces sons peuvent être faits de représentations communes (Michel Chion mentionne ici les pigeons de Paris) ou d'effets plus perceptifs. Ils sont des repaires spatiotemporels qui participent à la crédibilité de l'image et étoffent la bande-son.

L'utilisation du rendu multicanal (5.1 ou 7.1, le plus souvent) permet de laisser davantage d'espace aux ambiances, aux rumeurs (sur les canaux de côté ou arrière) tout en laissant les sons de premier plan au centre. C'est le cas notamment dans le film *Blade Runner* (1982) qui fournit une rumeur quasi continue de différentes textures et ambiances urbaines en support des dialogues [Chi94].

Les stratégies de sonorisation des villes au cinéma s'articulent le plus souvent autour du mélange de strates sonores de morphologies et de densités différentes, selon les objets sonifiés. Les représentations sonores symboliques ou exagérées priment sur la recherche de réalisme afin de magnifier la sensation ou l'effet qu'il s'agit de communiquer au spectateur. Les différents éléments sonores d'une scène peuvent ainsi se rattacher à un événement précis et individuel ou, à l'opposé, englober une masse d'éléments avec un objet sonore.

L'analyse des villes sonores au cinéma montre une distinction entre les sons relatifs aux actions et les sons relatifs aux ambiances. Il peut s'agir de deux niveaux de narrations différents, les arrière-plans sonores appartenant à la seconde catégorie. L'utilisation d'ambiances sonores participe de la crédibilité de la ville filmée. Les ambiances sont aussi le support des autres sons de l'environnement, permettant de conserver une unité de lieu et de temps lors des changements de plan.

Ce bref parcours des études sur la sonorisation des villes au cinéma donne un premier aperçu des stratégies mises en place pour recréer le cadre urbain. Dans les paragraphes suivants, nous prolongeons cette approche en nous intéressant aux villes virtuelles, principalement interactives.

2.3.2 Villes sonores interactives

L'étude des villes au cinéma permet d'aborder les stratégies de sonorisation linéaire des villes audiovisuelles. Cette thèse a pour objet de définir, spécifier et prototyper un outil de synthèse de l'arrière-plan urbain, utilisable dans des applications fortement interactives telles que le jeu vidéo. Nous souhaitons maintenant nous intéresser aux stratégies de sonorisation interactives dédiées aux villes virtuelles.

Nous tenterons donc de répondre aux questions suivantes : Quels sont les principaux mécanismes de l'écriture sonore interactive ? Que sont, d'un point de vue sonore, les villes virtuelles ? Quelle place occupent les arrière-plans sonores dans les villes virtuelles interactives et comment sont-ils conçus ?

Depuis les premiers modèles sonores du trafic [MH74], les efforts de recherche se sont affinés dans le domaine de la simulation acoustique dynamique [TFNC01] avec pour objectif une exhaustivité optimisée du rendu de scènes sonores complexes [MBT⁺07]. Si une des motivations premières à la virtualisation de la ville sonore est d'ordre écologique [KVB01], l'immersion recherchée dans l'interactivité ludique pousse également vers des modèles complexes [RSM⁺10]. Notre travail de thèse s'oriente davantage vers une recherche perceptive du son de la ville. Nous nous attardons donc davantage sur les techniques de sonorisation du monde du jeu vidéo que celles des modélisations physiques des environnements urbains.

La différence première entre le cinéma et un environnement virtuel est contenue dans l'interactivité. Là où le cinéma n'exige du spectateur que son attention, l'environnement virtuel requiert son action.

Il en résulte que l'illustration sonore d'un environnement virtuel ne se limite pas toujours à un accompagnement neutre mais peut aussi renvoyer des informations à l'utilisateur, en lien avec son action directe (signaux d'alerte, validation d'une action effectuée, etc.).

Karen Collins distingue trois niveaux d'interaction sonores possibles avec l'environnement virtuel [Col07] :

- Audio interactif : ce qui propose un retour direct à une action
- Audio adaptatif : ce qui réagit de façon appropriée et parfois anticipe les changements dans le jeu et/ou le gameplay
- Audio dynamique : comprend les deux points précédents et donc réagit aux changements dans l'environnement du gameplay ou en

réponse à une action du joueur

Une autre manière d'étudier le son des jeux vidéo est de répertorier les différentes catégories d'objets sonores. Stockburger propose une des premières classifications des différents types d'événements sonores présents dans un jeu vidéo [Sto03] :

- Voix (dialogues)
- Zone (ambiances)
- Partition (musiques)
- Effets (sons diégétiques du jeu)
- Sons des objets d'interface (sons des menus et plus généralement tout ce qui ne semble pas appartenir à la partie diégétique de l'environnement du jeu)

Dans une publication de 2011, Joergensen [Jør11] propose un tour d'horizon des différentes catégorisations de sons propres aux jeux vidéo et environnements virtuels.

Enfin, la spatialisation dynamique du son constitue une composante importante propre aux environnements interactifs. Elle peut servir à guider le joueur dans son action en fournissant, par exemple, des indications d'orientation pour situer la position d'un objet dans l'espace [Gon08].

2.3.2.1 Frontières diégétiques des environnements virtuels

Dans la théorie du cinéma, on appelle diégétiques les sons trouvant leur source dans l'environnement du film et non-diégétiques les sons provenant d'une source extérieure à l'environnement du film [Jør07]. Pour Collins [Col08], l'audio dynamique complexifie la séparation standard entre son diégétique et non-diégétique. Le fait que l'écouter est impliqué, par son action, dans le déroulement de la bande-son requiert un autre type de classification des relations son-image. S'il est possible de distinguer son diégétique et non-diégétique il faut également subdiviser ces grandes catégories entre son dynamique et son non-dynamique avant de distinguer les différents types d'activité dynamiques relevant de la diégèse et du joueur.

Ces différents niveaux rendent compte de la complexité de l'écriture sonore pour les médias interactifs en raison de la grande imprévisibilité des actions du joueur et du besoin de maintenir intelligibles les éléments sonores signifiants et immersifs. Cette frontière plus floue est le point d'origine d'une catégorie de sons spécifiques aux médias interactifs : les

sons trans-diégétiques [Jør07]. Pour Joergensen, les termes diégétiques et non-diégétiques ne prennent pas en compte les aspects fonctionnels du son dans les environnements interactifs, ce qui distingue ces derniers de l'environnement de la fiction. Certains sons entretiennent en effet une ambiguïté : ils apportent des informations d'ordre utilitaire au joueur tout en étant stylisés de manière à correspondre à l'univers du monde fictionnel. Ils sont à la fois un support pour le gameplay et participent de la sensation de présence dans l'univers du jeu. Les notions de diégétique et non-diégétique ne sont pas théorisées pour de telles situations et ne tiennent, historiquement, pas compte de l'intervention de l'écoutant dans le déroulement du média [Jør11]. Joergensen qualifie ces sons de trans-diégétiques, ils présentent la particularité de combiner les informations relatives au système du jeu avec l'environnement virtuel, créant ainsi une trame de référence qui, en plus de présenter une valeur d'utilité, appuie la sensation de présence dans l'univers du jeu.

Dans ce contexte, quelles sont les implications de l'intervention de l'écoutant pour les ambiances et les arrière-plans de l'environnement interactif ?

2.3.2.2 Gestion des arrière-plans sonores dans les jeux vidéo

Dans la production des jeux vidéo, les arrière-plans appartiennent à ce qu'il est également commun d'appeler les ambiances, c'est-à-dire les strates sonores sur lesquelles reposent les dialogues, effets et autres sons du jeu. Il s'agit, dans leur forme la plus simple, de boucles pré-enregistrées dont la durée peut varier de une à plusieurs minutes. Comme nous le verrons, les jeux recherchant une narration dynamique peuvent cependant requérir la mise en place d'un système plus complexe et moins linéaire.

Dans un article de 2007, Bridgett [Bri07b] attribue aux ambiances deux principales fonctions : elles offrent une continuité sonore au-delà de l'image et créent un liant pour les autres sons, assurant ainsi une sensation de continuité, d'homogénéité et d'unité à la scène. Si l'ambiance n'a pas lieu d'être réactive de façon synchrone avec la diégèse, elle n'est pas hermétique aux événements du jeu et se transforme en fonction de ceux-ci. Son effet est donc « subliminal », distillant de façon subtile des changements dans l'humeur ou la tension de l'environnement virtuel pour le joueur. En cela, les ambiances dynamiques peuvent recouvrer le

même rôle qu'une musique tout en restant au niveau du bruitage (apparition dans l'arrière-plan de sirènes ou de klaxons à la suite d'une action du joueur, par exemple). Il existe peu de littérature scientifique sur l'analyse de ces strates sonores et leurs modes de réalisation. Notre principale source d'information vient des studios de développement de jeux vidéo eux-mêmes, lorsque les créateurs sonores mettent à disposition un *post-mortem* expliquant leur méthodologie. Pour la conception des ambiances, Bridgett explique que l'enregistrement sur site d'ambiances est le meilleur moyen de collecter des données authentiques et d'avoir une trame du ressenti de l'environnement. Cette méthode peut s'appliquer aussi à des environnements fictifs en enregistrant un environnement réel approchant les caractéristiques recherchées. Les enregistrements audio ne peuvent rendre compte qu'approximativement de l'atmosphère d'un lieu et il convient de collecter aussi des données visuelles (photographies, vidéos) afin d'avoir le matériel le plus riche possible.

Quel est le rôle de l'arrière-plan dans la construction de l'immersion et de la narration dans les jeux vidéos ? Quelles sont les stratégies de réalisation des arrière-plans ? Quels sont les enjeux techniques de la réalisation des arrière-plans ? Les trois paragraphes suivants sont consacrés à l'analyse de trois jeux vidéo se déroulant en milieu urbain ou dans un environnement d'inspiration urbaine.

Halo Combat Evolved (Bungie, 2001)

Halo est un jeu de tir à la première personne se déroulant dans un univers futuriste avec des ambiances sonores complexes réparties sur plusieurs pistes audio [O'D02]. Ainsi, la sonorisation de l'ambiance d'une plage est composée d'ensembles de boucles audio dédiées à des textures environnementales (mer, vent) sans spatialisation particulière, agrémentées de sons déclenchés sporadiquement (oiseaux, gravas) et répartis aléatoirement dans l'espace sonore 3D.

Les musiques sont présentes très ponctuellement et appuient la narration avec des empreintes émotionnelles. O'Donnell explique la distinction et la complémentarité des ambiances et de la musique en attribuant un caractère immersif aux premières et la recherche d'impact émotionnel à la seconde. La musique est faite essentiellement de segments pré-enregistrés et divisés en trois parties : un segment d'in-

troduction, un segment mis en boucle (avec une version alternative) et un segment de sortie (avec une version alternative aussi).

La musique et les ambiances occupent une place à part dans la sonorisation de Halo. Ces deux strates exploitent la sonorisation multicanale du jeu (5.1) mais ne sont pas disposées selon la géométrie : la musique et les ambiances sont en stéréo, cinquante pour-cent du signal est renvoyé vers les enceintes arrières et le caisson de fréquences graves.

Si les principes mis en œuvre pour la production sonore de Halo semblent simples de par le recours quasi exclusif à des boucles pré-enregistrées, l'efficacité de l'ensemble tient principalement dans l'attention apportée au mixage et à la production des boucles en prévision de leurs interactions avec les autres sons du jeu. O'Donnel précise que le mixage dynamique de certains sons d'ambiance aurait pu bénéficier de variations liées à la position du joueur et pas seulement au temps.

L'exemple de *Halo* fait ressortir l'utilisation des arrière-plans sonores à des fins immersives. En cela, ils sont bien dans la continuité des arrière-plans sonores au cinéma, servant de support aux autres sons de la scène. Cependant, un degré supplémentaire de complexité est ici ajouté avec certains éléments de l'arrière-plan étant déclenchés aléatoirement et disposés dans la stéréo.

Bioshock (2K Games, 2007)

La frontière plus flottante entre sons diégétiques et non-diégétiques ainsi que la possibilité d'utiliser des sons trans-diégétiques peut participer de la stratégie de sonorisation. C'est le cas de certains jeux de type *survival horror*. Le son possède un rôle central dans ce type de jeux avec pour objectif d'accentuer la peur et la tension pour le joueur. Pour Kromand [Kro08], la stratégie pour parvenir à un tel résultat consiste principalement à brouiller la frontière entre les sons diégétiques et non-diégétiques, avec un recours important aux sons trans-diégétiques.

Cette stratégie se base sur la connaissance implicite du joueur de l'environnement sonore dans lequel il évolue et plus précisément de sa capacité à discerner les objets sonores signifiants du reste du paysage sonore du jeu. Kromand analyse l'ambiance sonore du jeu de tir Bioshock, se déroulant dans une ville imaginaire et abandonnée de presque tous ses habitants, Rapture (Fig. 2.14).

Kromand donne une interprétation du rôle de l'arrière-plan sonore en distinguant quatre strates qui le composent :

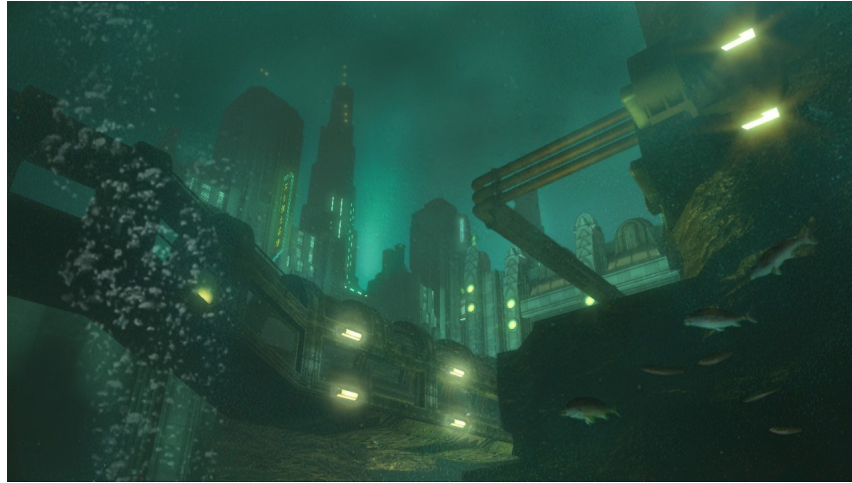


FIGURE 2.14 – Image extraite du jeu vidéo *Bioshock* (2K Games, 2007), vue du principal environnement du jeu : la ville *Rapture*

- *Drones continus évoquant des machines.* Ces drones varient d'un endroit à un autre dans le parcours du jeu. Ils participent de l'ambiance du jeu telle que la définit Jorgensen, c'est-à-dire comme une atmosphère générale liée à la zone où se trouve actuellement le joueur.
- *Un battement dans les fréquences graves* est également présent, clairement audible à de forts niveaux seulement, ce battement aide à combler les silences et sert de support aux autres sons non-diégétiques.
- *Strates musicales en forme de crescendos.* Les crescendos sont connus pour leur utilisation dans les films d'horreur en anticipation d'un événement qui va causer la peur chez le spectateur. Les mêmes effets se retrouvent dans les ambiances de Bioshock mais sont déclenchés sur une base temporelle et non en lien avec la narration. Kromand argumente qu'il s'agit d'un effet reposant sur l'affordance pour accroître la tension chez le joueur en suscitant chez lui l'anticipation d'une rencontre avec un ennemi⁹.

9. Notons également l'utilisation des effets sonores musicaux dans le jeu Dead Space [Nap08]. La proximité d'un ennemi déclenche le mixage dynamique de strates musicales dissonantes (principalement à bases de clusters, glissandos et autres techniques inspirées de la musique contemporaine) : les strates sonores sont audibles avant que l'ennemi ne soit visible et leur intensité s'accroît avec la proximité de ce

- *Sons aléatoires identifiables sans être directement reliés à la scène* (craquements métalliques, gouttes ou ruissellements). Grimshaw [Kro08] considère ces sons, proches du diégétiques, comme des *sensory fillers*. Ils sont des perturbateurs pour le joueur : en remplissant le paysage sonore, ils surchargent d'informations la cognition du joueur et masquent les sons de potentiels ennemis ; la capacité du joueur à lire au travers de la trame sonore pour anticiper une rencontre avec un ennemi est ainsi mise à l'épreuve. Une relation double se crée : d'un côté la position spatiale des ennemis peut être déduite par leurs émissions sonores mais d'un autre côté les ambiances brouillent la lisibilité de la scène entre les sons diégétiques et non-diégétiques.

Les arrière-plans sonores du jeu vidéo Bioshock apparaissent plus complexes dans leur composition et leur interaction que des pistes d'accompagnement passif. L'arrière-plan est ici un mixage dynamique de plusieurs sous-ensembles sonores. L'état de l'environnement ainsi que la position du joueur dans l'environnement sont pris en compte afin de produire un rendu sonore dynamique et immersif.

Kromand procède à l'analyse de deux autres jeux appartenant au genre *survival horror*, F.E.A.R. et Silent Hill 2, exploitant des codes de sonorisation de l'arrière-plan très proches de ceux décrits ici. Les difficultés du joueur à discerner les sons diégétiques des sons non-diégétiques créent ce que Kromand appelle un effondrement de la barrière diégétique. Ces difficultés sont volontairement induites par les concepteurs du jeu afin de renforcer l'immersion en augmentant la tension dans le contexte du jeu. Elles sont renforcées par l'absence de tonalité des ambiances et l'incorporation d'éléments sonores susceptibles d'être facilement pris pour diégétiques.

Prototype (Radical Entertainment, 2009)

Prototype est un jeu d'action/aventure à la troisième personne. Il se déroule dans un univers de type « bac à sable » c'est-à-dire que le joueur possède une liberté presque totale de déplacement dans l'envi-

dernier [Mac09]. Cette strate agit comme un signal d'alarme pour un danger mais n'indique volontairement pas la proximité temporelle et/ou spatiale précise de ce danger.

ronnement du jeu. Morgan¹⁰ propose un *post-mortem* de la réalisation sonore du jeu [Mor09].

La charte sonore du jeu proposait de reconstituer la sensation sonore propre à la ville de New York, sans pour autant aller vers un résultat exhaustif. Un accent particulier est mis sur l'aspect dynamique des ambiances qui doivent réagir aux différents états du jeu ainsi qu'aux actions du joueur. Le jeu se déroulant dans un environnement urbain très habité, un des enjeux de la sonorisation est de réussir la gestion de centaines d'agents simultanément à l'écran.

La première étape dans la construction des ambiances était d'effectuer une vingtaine d'heures d'enregistrements à New York. Ces enregistrements stéréo ont eu lieu dans les conditions les plus diverses possibles et notamment depuis des points très différents : ruelles, carrefours, sommets d'immeubles ou de parkings.



FIGURE 2.15 – Image extraite du jeu vidéo *Prototype* (Radical Entertainment, 2009)

Les enregistrements qui se sont avérés être les plus utiles pour reconstituer des ambiances étaient ceux effectués depuis des points éloignés de toute source sonore individuelle clairement identifiables (toits, parcs, ruelles). Les plus silencieux d'entre eux permettent de révéler le bourdon ou la note fondamentale de la ville (« *keynote* »), ce sont

10. Scott Morgan était directeur audio pour le développement du projet Prototype

donc ces enregistrements qui ont servi de base à la reconstruction des ambiances finales en quadriphonie.

L'ensemble du paysage sonore du jeu se divise en trois tiers, ou trois plans acoustiques destinés à créer une sensation de perspective :

- Un *arrière-plan* composé de pistes d'ambiance en quadriphonie réalisé en combinant deux enregistrements stéréo différents pour accentuer les sensations de densité et d'espace.
- Un *plan moyen* destiné à sonoriser les groupes d'objets du jeu, tels que des piétons ou des véhicules ; il s'agit d'un plan central de la sonorisation du jeu en raison de sa réactivité aux événements du jeu.
- Un *premier plan* composé de sources et de sons individuels dont l'origine est directement observable tels que du dialogue ambiant, des klaxons, des tirs ou des crissements de pneus

Le mélange des trois strates permet de maintenir une cohérence entre tous les sons et empêche certains de ressortir abruptement du mixage. D'autre part, l'arrière-plan et le plan-moyen fournissent un support suffisamment dense pour réduire la quantité de sons individuels au premier plan. Le jeu met régulièrement en scène des centaines d'objets simultanément à l'écran, cette stratégie permet d'économiser des ressources pour les éléments cruciaux du *gameplay*.

La sonorisation de l'arrière-plan et des groupes (plan moyen) est assurée par un fichier multicanal de dix-huit pistes mixées dynamiquement selon la position de l'écouter dans l'environnement (principalement sa hauteur par rapport au sol). Les différentes densités de personnages à proximité du joueur, ainsi que leurs états dans le jeu, sont également utilisées comme des variables d'entrée pour le mixage dynamique. Ces deux strates sont ensuite traitées avec une réverbération et un filtrage procéduraux qui assurent la cohérence acoustique avec l'espace du jeu. Le mixage dynamique est censé rendre compte de l'état émotionnel de la scène autour du joueur en soulignant, par exemple, les accès de panique qui traversent la foule en cas d'affrontement entre le joueur et les personnages du jeu.

L'analyse de la dynamique des arrière-plans fait ressortir trois principaux points d'entrée : la densité de la scène autour du joueur, l'état de la scène et l'altitude du joueur dans l'environnement. Ce sont les trois principaux facteurs de mixage de l'arrière-plan virtuel.

Un outil audio propriétaire, AudioBuilder, assure l'interface entre le code du jeu et le mixage audio. Cet outil, inspiré des environnements

de programmation graphique Reaktor et Max/Msp, permet de régler les paramètres du mixage interactif selon les différentes entrées des systèmes du jeu.

Si ce système fonctionne bien dans l'ensemble, Morgan souligne tout de même quelques critiques. Notons surtout que le mixage dynamique des groupes ne permet pas toujours d'obtenir une réactivité suffisante en raison de la durée des fondus enchaînés entre les pistes correspondant aux différents états. Ces fondus enchaînés sont nécessaires afin d'assurer la fluidité et le naturel dans les transitions. Ensuite, le recours à des boucles audio sous-entend la possibilité de répétition et donc de prédictibilité, ce qui n'est pas souhaitable dans des ambiances. Le principal point positif est la nature dynamique du système qui permet de gérer, avec relativement peu de ressources, toutes les configurations de scène du jeu.

2.3.2.3 Conclusion

Ces différentes analyses montrent l'importance du matériau sonore de départ dans la construction des ambiances, notamment l'enregistrement d'ambiances les plus proches possibles de celles recherchées dans l'environnement virtuel. Cette stratégie peut différer des méthodes de sonorisation pour les autres objets sonores d'un jeu, la source sonore réelle ne correspondant pas systématiquement à l'objet virtuel qu'elle sonorise. L'étape de l'édition des enregistrements reste cependant d'égale importance. Rien ne doit trop ressortir des ambiances éditées mais il faut suffisamment de matière pour que le fichier sonore ait une identité [The12]. Parvenir à ce résultat requiert souvent un travail d'édition précis.

Nous observons aussi que les ambiances des environnements virtuels peuvent être constituées de plusieurs objets sonores superposés, multipliant ainsi les niveaux de lecture. Si certains de ces objets peuvent être constants dans tout l'environnement, d'autres sont liés à des zones géographiques délimitées. Si certaines strates de l'arrière-plan définissent l'atmosphère de l'environnement à des fins strictement immersives, d'autres permettent d'anticiper des événements. Certaines strates de l'arrière-plan peuvent agir comme un liant (ou être détournées en effet de masquage) pour des sons passant au premier plan, leur conception ne peut se faire sans tenir compte de la présence des autres sons de la scène au risque de masquer certains éléments ou d'en

mettre d'autres trop en avant. Le pré-mixage des ambiances est donc un élément fondamental de leur intégration pour leur efficacité et leur crédibilité [The12].

Du point de vue de l'écriture sonore de la scène, les arrière-plans peuvent se situer à la charnière entre le diégétique et le non-diégétique, selon le type de données qu'ils apportent au joueur. Dans la construction d'une ambiance, il est important de veiller à l'absence de répétitions perceptibles : la répétition et la prévisibilité nuisent aux qualités immersives des strates d'ambiance. Le travail d'édition, en retirant des éléments trop distincts de l'enregistrement, sert en partie à prévenir les répétitions au profit de l'uniformité.

D'autre part, l'exemple de Prototype révèle des méthodes d'enregistrements des ambiances lointaines et des arrière-plans urbains, depuis des points en hauteur ou des lieux éloignés de sources directes (comme des parcs).

Le mixage dynamique est un élément central dans l'efficacité des ambiances [Bri07a], il permet de séparer les plans sonores et de contrôler leur équilibrage respectif. Si les réglages du mixage dynamique jouent un rôle primordial, les jeux récents n'intègrent pas nécessairement des techniques procédurales dans la génération des ambiances. Ainsi, le jeu L.A Noire (Rockstar Games, 2011) comprend une centaine d'ambiances différentes chacune sonorisée par une boucle d'une durée d'une minute et demie à trois minutes [The12].

Les trois cas étudiés font ressortir des constantes dans la gestion des arrière-plans sonores dynamiques :

- Les arrière-plans, musicaux ou non, réagissent aux actions du joueur ainsi qu'à l'état général de la scène.
- Ils servent de support aux autres sons de l'environnement, ils agissent comme un liant assurant à la fois la cohésion spatiale des différents éléments de la scène et proposant une continuité hors-champs du paysage sonore.
- La sensation véhiculée par les ambiances prime sur leur réalisme, même si elles sont construites à partir de matériaux réels (comme pour Prototype), leur assemblage obéit davantage à des pré-requis perceptifs qu'à une recherche de réalisme acoustique.
- Les arrière-plans occupent l'ensemble de l'espace acoustique du jeu (stéréo, quadriphonie, 5.1, etc.).

De façon plus spécifique, quels sont les principaux éléments constitutifs des arrière-plans sonores dynamiques ?

- Des sons environnementaux statiques provenant par exemple de *field recordings*.
- Des sons environnementaux ponctuels (klaxons, sirènes, oiseaux, etc.).
- Des sons à caractère émotionnel : il peut s'agir de strates musicales ou d'effets sonores, conçus pour entretenir une mise en tension, par exemple.

Quelles sont les principales fonctions des arrière-plans dans les environnements virtuels interactifs ?

- L'immersion, en proposant des strates de sons environnementaux fortement spatialisées avec des composantes d'aléatoire (afin d'apporter de la diversité et d'éviter les sensations de répétition)
- Comblent les silences. Les actions du joueur étant le plus souvent imprévisibles au moment de la conception du jeu, les ambiances et arrière-plans peuvent garantir la présence d'éléments sonores de l'univers du jeu, y compris dans les phases d'inaction, afin de ne pas rompre l'immersion.
- Lier et/ou masquer les autres sources sonores. Dans le mixage dynamique, l'arrière-plan sonore devient une strate qui se superpose aux autres éléments du jeu. Le choix des éléments qui le composent et les variations de volume permettent de créer une sensation de continuité entre l'arrière-plan et les autres objets sonores de la scène. Cette continuité peut servir de liant pour tous les sons de la scène, ou de masque pour certains d'entre eux, selon les intentions du concepteur sonore. L'effet de masquage peut également être mis à profit pour réduire la quantité de sons lus en simultané.

Et finalement, quels éléments de l'environnement sont dynamiquement pris en compte pour faire varier les arrière-plans sonores ?

- La position du joueur dans l'environnement. Les arrière-plans sont souvent composés pour chaque zone du jeu, la position du

joueur détermine quel arrière-plan doit être entendu. L'un des principaux enjeux lors de la création des ambiances dynamiques d'un jeu est de penser les transitions entre les différentes ambiances.

- Le temps. Il peut servir de base pour le déclenchement de sons aléatoirement ou créer des variations automatiques afin d'éviter la sensation de répétition.
- La proximité du joueur à certains objets de l'environnement.

Cette section nous a permis d'aborder les stratégies de sonorisation de la ville audiovisuelle au cinéma et dans les jeux vidéo. Les exemples étudiés font ressortir certaines constantes et mécaniques dans la création et l'utilisation d'arrière-plans sonores dynamiques.

Dans la section suivante, nous nous intéressons aux outils de création sonore dédiés aux environnements virtuels ou adaptés aux villes virtuelles. Et nous verrons dans quelle mesure ces outils permettent de développer des arrière-plans dynamiques.

2.4 Outils de sonorisation des villes virtuelles

La nature des villes virtuelles telles que traitées dans cette thèse requiert des outils de sonorisation spécifiques. L'interactivité implique de faire communiquer la partie sonore avec d'autres sections du média (gestion de la physique, intelligence artificielle, position de la caméra,...). Les outils mis en pratique pour la sonorisation de villes virtuelles font souvent l'objet d'un compromis entre possibilités d'interaction et ressources disponibles.

La section précédente a mis en lumière des stratégies de sonorisation propres aux environnements virtuels urbains, ou d'inspiration urbaine. Cette section entre davantage dans le détail des outils disponibles pour la sonorisation d'environnements complexes et propose de les comparer succinctement afin de mieux cerner leurs possibilités.

Les outils aujourd'hui à disposition des concepteurs et designers sonores proposent des approches parfois radicalement différentes les unes des autres. Nous procéderons dans un premier temps à un comparatif entre différentes méthodes de synthèses utilisées dans la réalisation de textures complexes. Ensuite, nous aborderons spécifiquement les outils de sonorisation dédiés à la production de jeux vidéo. Nous pensons que la mise en regard de ces deux axes devrait nous permettre de mieux comprendre les enjeux techniques liés à la synthèse temps-réel d'arrière-plans sonores urbains.

2.4.1 Synthèse en temps réel d'ambiances urbaines

Dans cette section, nous faisons un comparatif des différentes méthodes de synthèse temps-réel préexistantes qui sont à disposition pour réaliser un synthétiseur d'ambiances urbaines. Ces différentes méthodes sont comparées au regard de leurs possibilités en matière de génération de textures.

La complexité des scènes urbaines requiert une approche *environnementale* [MC09] de leur synthèse sonore, c'est-à-dire qu'il s'agit de combiner différents types de sources sonores situées au premier et à l'arrière-plan : des textures ambiantes, des événements sonores contenant plus ou moins de bruit ou encore des voix. Afin de répondre à la diversité des sources incluse, différents paradigmes de d'analyse et de synthèse peuvent être utilisés en parallèle. Cette idée est prolongée par Andy Farnell avec son ouvrage *Designing sound* [Far10] dans lequel il

préconise d'étudier les mécanismes de la source à synthétiser afin de déterminer la méthode de synthèse la plus efficace.

Il existe un certain nombre de classifications des différentes méthodes de synthèse sonore (par exemple dans [Sch11] et [MC09]). La présentation que nous proposons ici s'organise autour de l'intégration des méthodes de synthèse sonore dans les moteurs audio dédiés aux applications interactives et aux jeux vidéo. D'autre part, si toutes les méthodes présentées dans les paragraphes suivants ne sont pas directement pertinentes pour la synthèse sonore d'arrière-plans urbain, nous pensons que les paradigmes qu'elles mettent en œuvre constituent souvent des points de réflexion intéressants dans la constitution d'un nouvel outil.

Nous présentons dans un premier temps des méthodes pouvant être appliquées au rendu d'environnements sonores complexes. Nous nous intéressons ensuite aux méthodes de synthèses par modulation de fréquence et par tables d'ondes qui, si elles sont moins présentes dans les moteurs audio d'environnements virtuels, n'en présentent pas moins un certain attrait au regard de leurs possibilités en terme de design sonore. Enfin, nous présentons plus spécifiquement la synthèse granulaire avant de proposer un bilan comparatif de ces différentes méthodes.

2.4.1.1 Synthèses environnementales

Les méthodes de synthèses environnementales que nous présentons brièvement dans les paragraphes suivants consistent à approcher au plus près les caractéristiques préalablement analysées du son avant de le modéliser. Elles procèdent d'abord à une analyse des matériaux de la source ou d'un enregistrement sonore de celle-ci avant d'en proposer un modèle manipulable dans le domaine temporel et/ou fréquentiel. Nous présentons ici trois méthodes principales, la modélisation physique, les modèles physiquement informés et la synthèse par ondelettes.

La synthèse par modélisation physique consiste à générer une forme d'onde selon un modèle mathématique des propriétés physiques de la source ainsi que la manière dont cette dernière est excitée (frappée, frottée, soufflée). Ce type de synthèse est souvent utilisé afin de reproduire des instruments de musique [ICD97] ou pour imiter des matériaux tels que le métal, le bois ou le verre [VdDKP01]. L'approche modale est une des plus utilisées, elle repose sur la description d'objets mis en vibration, le plus souvent des cordes ou des plaques, pouvant être com-

binés entre eux pour accroître la complexité de la géométrie simulée [VdDP03]. La modélisation est plus efficace que dans le domaine temporel que dans le domaine fréquentiel [BDT⁺08], elle permet de calculer simultanément un grand nombre de sources, principalement pour des sons de frottements ou d'impacts. Ces descriptions, au premier abord rudimentaires, peuvent mener à la description d'environnements complexes et de multiples formes d'interactions entre les éléments qui le composent [Men10].

Dans une approche plus fine, ces méthodes de synthèse sont souvent dites *physiquement informées* (*physically-informed*) c'est-à-dire qu'elles utilisent une analyse du matériau source à synthétiser pour contrôler les comportements du modèle physique [Coo97]. La compagnie AudioGaming a récemment sorti un *plug-in* de design sonore dédié aux sons environnementaux de vents et de pluie, *AudioWeather*¹¹, qui utilise une méthode physiquement informée. Les paramètres haut-niveau d'édition de *AudioWeather* sont construits sur des critères perceptifs et affectent chacun plusieurs paramètres bas-niveau de la synthèse [Mal11]. Un autre exemple est le synthétiseur environnemental proposé par Charles Verron [Ver10] qui se base sur des modèles issus de la taxonomie de Gaver [Gav93] (aérodynamiques, liquides et solides) tout en utilisant un mélange de synthèse additive et de bruit pour reconstruire le signal analysé. Un des avantages mis en avant par Verron est la possibilité de contrôler finement la spatialisation de chacun des éléments de la synthèse au moment de la synthèse [RD92].

Enfin, la synthèse sonore par ondelettes [KM88] est une méthode qui consiste à analyser et recomposer un signal [Sch11] et se distinguant de la transformée de Fourier [Gra95]. La synthèse par ondelettes se déroule en plusieurs phases [MC05]. D'abord une phase d'analyse qui se fait par la décomposition du signal en ondelettes avec la méthode DWT (*Discrete Wavelet Transform*), un modèle du son original est extrait en utilisant les coefficients de décomposition, le son original peut être parfaitement reconstruit. La décomposition est une base pour générer un certain nombre de paramètres qui permettent, par leur manipulation de produire des sons similaires au modèle. Ce type de synthèse fonctionne bien avec des textures stochastiques et non stationnaires [Fin09]. Un exemple d'utilisation de cette synthèse avec Pure Data est

11. <http://www.audiogaming.net/audiowind-audiorain>

décrit dans [DP06].

L'utilisation de ces trois méthodes présente toutefois certaines limitations. Le temps de calcul nécessaire à l'utilisation de ces méthodes peut s'avérer très important. D'autre part, la démarche d'analyse-resynthèse peut être coûteuse en temps de production pour parvenir à la resynthèse d'une texture prédéfinie. Cependant, une fois le processus d'analyse achevé, ces méthodes permettent d'obtenir des variations sonores intéressantes et efficaces de sons environnementaux parfois complexes.

2.4.1.2 FM et tables d'ondes

Les méthodes de synthèse suivantes n'ont pas été directement conçues pour l'émulation de sons environnementaux. Cependant, leur versatilité dans les timbres produits en font des méthodes potentiellement intéressantes pour la synthèse de timbres complexes.

La synthèse par modulation de fréquences (*Frequency Modulation* ou plus simplement *FM*) [Cho73] connaît ses premières implémentations commerciales auprès de Yamaha qui a popularisé cette méthode, notamment avec le synthétiseur *DX7*. Ce type de synthèse se prête bien aux sons de nature percussive et peut être utilisé également pour reproduire des voix [Cho89]. Sa principale limitation au regard de l'interactivité reste cependant liée à son architecture. La complexité et la précision des timbres obtenus sont en grande partie liées au nombre d'opérateurs disponibles (soit d'oscillateurs) et aux possibilités d'interactions entre eux. Or, plus le nombre d'opérateurs est important et plus le coût de calcul est élevé. Des expérimentations ont cependant lieu pour une utilisation interactive de la synthèse FM [LGMV07] mais principalement à des fins musicales.

La synthèse par tables d'ondes (*wavetable synthesis* [And79]), fut développée par Wolfgang Palm pour PPG (Palm Products GmbH) et implémentée principalement dans les synthétiseurs des marques PPG (Fig. 2.17) puis Waldorf Music. La synthèse par table d'ondes est basée sur l'utilisation de formes d'ondes périodiques d'un cycle, générées en temps réel par de la synthèse additive contrainte [BJ96]. La table d'ondes peut être modifiée et parcourue à l'aide de contrôles dédiés, ce qui permet de faire varier le timbre produit en temps-réel. La qualité d'une table d'ondes réside dans le choix des formes d'onde de base et



FIGURE 2.16 – Photographie du panneau avant du module de synthèse par table d'ondes *E350 Morphing Terrarium* (Synthesis Technology, 2012), ce module numérique est destiné aux synthétiseurs modulaires de format *Eurorack*.

dans les algorithmes d'interpolation entre elles. Très prisée pour ses capacités à produire des timbres complexes, évolutifs et abstraits, la synthèse par table d'ondes est à ce jour utilisée dans de nombreux modules de synthèse virtuels ou physiques (Native Instruments *Massive*, Waldorf *Largo*, Synthesis Technology *E350*,...). La synthèse par table d'ondes est souvent confondue avec la synthèse à base d'oscillateurs de type *table lookup* (utilisant des formes d'ondes d'un cycle stockées en mémoire) ou la synthèse dite *PCM* (*Pulse-Code Modulation*) utilisant des échantillons pré-enregistrés et stockés en mémoire.

Ces méthodes, très populaires dans les années 1980 et 1990 (notamment la synthèse FM), sont associées à des sonorités particulières et qui ne sont pas systématiquement recherchées au moment de l'élaboration des jeux actuels. Les synthétiseurs *vintage* font cependant partie inté-



FIGURE 2.17 – Photographie d'un système PPG *Wave 2* complet (1981) incluant le moniteur *Waveterm* de visualisation de tables d'ondes.

grante du matériel utilisé par les designers sonores pour créer des éléments sonores (par exemple dans le cas du jeu *Dead Space*[Nap08][Mac09]) qui sont ensuite intégrés à l'univers du jeu. Les ressources mises à disposition pour la création des jeux vidéo augmentant régulièrement, les synthèses FM et par table d'ondes seront peut-être plus largement mises à disposition pour la création des bande-sons, notamment en raison de leurs versatilité pour les timbres musicaux. Cependant, des évolutions de ces modèles, comme la synthèse par *wavescanning* [VMS01] (qui peut être vue comme un développement des *wavetables* dans la mesure où il s'agit de rendre dynamiques les tables d'ondes [BSF00] afin de reproduire plus finement des timbres complexes), laissent entrevoir des possibilités de design sonore auparavant inédites. Nous notons cependant avec intérêt le concept de la table d'ondes offrant une grande palette de timbres parcourue à l'aide d'un seul paramètre.

2.4.1.3 Synthèse granulaire

La synthèse granulaire se trouve plus fréquemment employée en temps-réel dans la sonorisation d'environnements virtuels que les synthèses FM et par table d'ondes. La synthèse granulaire recouvre di-

verses méthodes complexes de lecture de fichiers sonores pré-enregistrés. Contrairement aux méthodes décrites dans les paragraphes précédents, elle ne consiste pas, initialement, à produire un signal à l'aide de paramètres descriptifs mais à exploiter les possibilités d'édition en temps-réel et de lecture d'un fichier sonore préexistant.

Principe général

La granulation d'un signal consiste à « produire des textures sonores en manipulant des grains de son (échantillons plus ou moins courts) » [PS05]. On peut trouver les prémices modernes de la synthèse granulaire dans un article du physicien Denis Gabor, publié en 1947 par *Nature*, dont l'anticipation de l'application de la théorie quantique aux signaux acoustiques [Gab47] fut ensuite développée par Wiener [Wie64]. Iannis Xenakis a théorisé une méthode de composition reposant sur l'utilisation de grains sonores [Xen63]. Barry Truax [Tru90] développe ensuite cette méthode pour la composition d'environnements sonores [Tru92]. Aujourd'hui, l'ouvrage *Microsounds* [Roa04] de Curtis Roads offre un panel très complet des possibilités de la synthèse granulaire, cette lecture peut être complétée par le document de thèse « Les techniques granulaires dans la synthèse sonore » de Manuel Rocha Iturbide [Roc99].

Il n'existe donc pas une méthode de synthèse granulaire mais davantage un principe de granulation sonore décliné en une variété d'applications. Le choix des paramètres de la méthode est le plus souvent influencé par l'anticipation du résultat sonore souhaité. La synthèse granulaire permet d'obtenir des sonorités proches (des variations) du fichier d'origine ou bien, à l'opposé, de créer des textures entièrement nouvelles et souvent complexes de par l'entrelacement des grains lus selon des paramètres variant en temps-réel.

Ce principe général se décline en deux sous-ensembles couvrant des finalités distinctes fréquentielle et temporelle [Roc99]. Le premier sous-ensemble concerne les méthode d'analyse-resynthèse tandis que le second se concentre sur le contrôle total des grains. Les méthodes d'analyse-resynthèse sont plus adaptées à la création de « micro-structures », soit de timbres, tandis que les méthodes ne reposant pas sur une analyse sont plus adaptées à la création de « macro-structures » sonores [Roc99].

Le principe est de segmenter le fichier sonore en grains et de lire ces grains en les mélangeant entre eux. La taille des grains varie le

plus souvent entre 50 et 100 millisecondes, elle peut être moindre mais peut parfois aussi atteindre plusieurs secondes [KT98]. À chaque grain est appliquée une enveloppe d'amplitude (ou fenêtre) afin d'éviter les artefacts liés à la rupture de la forme d'onde aux extrémités des grains. La forme de cette enveloppe est variable (Hanning, rectangulaire, linéaire, exponentielle,...) et contribue grandement au rendu sonore dans la mesure où elle donne la forme dynamique des grains dans le temps et définit le type de recouvrement entre les grains. La constitution des grains (position dans le fichier d'origine et taille), l'ordre et la vitesse à laquelle ils sont joués, la fenêtre d'amplitude qui leur est appliquée, leurs relations temporelles, leur mélange ou encore leur position spatiale sont autant de paramètres prédéfinis et/ou éditables [Roa04] dans une méthode de synthèse granulaire.

Nous reprenons la distinction entre les re-synthèses granulaires dites *synchrone*, *asynchrone* et *quasi-synchrone*. La synthèse granulaire *synchrone* se caractérise par la régularité des intervalles entre chacun des grains ([Roa04] p.93). Les applications musicales de la re-synthèse granulaire bénéficient de cette régularité tant pour l'élaboration de structures rythmiques que pour la génération de timbres. La synthèse granulaire *asynchrone* se caractérise par « une grande liberté dans la manipulation des paramètres du grain » [PS05] en temps réel. Ces paramètres incluent la taille de la fenêtre (soit la durée du grain), le type de la fenêtre (soit la forme d'onde de l'enveloppe dynamique appliquée à chaque grain), les caractéristiques temporelles du fenêtrage, la densité de la granulation. La forme asynchrone convient plus particulièrement à la génération de nuages de sons basés sur des algorithmes stochastiques ou chaotiques ([Roa04] p.96). La synthèse granulaire dite *quasi-synchrone* inclue des déviations aléatoires plus ou moins importantes pouvant évoquer les deux formes précédemment présentées, selon la quantité de déviation appliquée ([Roa04] p.93).

Principales applications

Il serait difficile de dresser une liste exhaustive des applications de la synthèse granulaire tant celles-ci sont diverses. Nous pouvons en revanche mettre en avant des effets possibles que nous croyons particulièrement pertinents dans le contexte de cette thèse. Rocha [Roc99] distingue dans les travaux de Roads trois approches que nous détaillons dans les paragraphes suivants.

Tout d'abord, les « effets changeants de temps/taux » qui permettent de faire varier la durée d'un son sans en changer la hauteur musicale perçue et inversement. Ces techniques sont aujourd'hui communément rassemblées sous le terme *time warping* et font référence, respectivement, au *time-stretching* et *pitch-shifting*. La mise en place de ces méthodes repose sur le choix des grains d'un son à lire et/ou leur étirement compensé par une variation du taux d'échantillonnage.

Ensuite, la « granulation temporelle asynchrone » qui est appliquée à un fichier sonore stocké en mémoire et dont les grains sont extraits de diverses manières : séquences continues ou discontinues, dans le sens de lecture du fichier ou dans le sens inverse, de manière aléatoire... Plusieurs fichiers sonores peuvent ainsi servir de source pour créer des nuages de « texture hybride ». Les caractéristiques du nuage ainsi créé dépendent essentiellement des fichiers source utilisés et des relations fréquentielles, musicales, sémantiques, ou encore morphologiques qui pré-existent entre eux. La transposition harmonique du fichier source est également possible en changeant son taux d'échantillonnage. La thèse de Rocha présente un panel des larges applications de la granulation temporelle asynchrone [Roc99]. Nous avons mentionné plus haut la variabilité de la taille des grains. Si cette variation peut avoir lieu d'un grain à un autre, par exemple, il convient de respecter une durée minimale ([Roc99] p.100) afin de préserver les « caractéristiques intrinsèques de la source sonore » :

Par exemple, pour des fichiers sonores de voix parlée, il faut avoir des durées de grain assez longues (entre 30 et 50 msec) pour pouvoir identifier la source (Roads, 1991). Autrement, si on utilise des petits grains de 10 msec ou moins, la source devient de moins en moins perceptible, et les effets sonores de modulation vont prédominer sur l'échantillon originel.

Enfin, le troisième aspect principal que nous retiendrons ici concerne la spatialisation, soit la distribution dans plusieurs canaux audio des grains audio. Plusieurs manières de procéder sont possibles (voir la thèse de Rocha [Roc99] p.87) et souvent relatives à la densité du nuage de grains créé par la resynthèse. Nous noterons principalement ici l'exemple d'application avec le synthétiseur *AmbiGrainer* développé par Mariette [Mar09] donne un aperçu des possibilités effectives

de la synthèse granulaire, notamment en ce qui concerne la spatialisation des grains en ambisonie.

Granulation et médias interactifs temps-réel

Dans le cadre des médias interactifs, la synthèse granulaire offre un compromis. Dans un premier temps, l'enregistrement de matériaux sonores bruts est un point de départ valable pour disposer de timbres riches et uniques. Dans un deuxième temps, la nature même de la lecture d'échantillons de la synthèse granulaire est une manière de contourner le recours aux fichiers mis en boucle. Enfin, une série de traitements appliqués sur les grains permet un apport conséquent en terme de design sonore. Ces traitements peuvent concerner le timbre des textures et la spatialisation des grains. Le concept de granulation a connu, ces dernières années, une évolution vers la synthèse dite *concaténative par corpus* [Sch06] dont le principe est de trier les grains selon diverses propriétés extraites par des descripteurs. Il devient donc théoriquement possible d'organiser automatiquement le contenu d'un fichier audio et de naviguer à l'intérieur de celui-ci selon des paramètres haut-niveau. Ce plus grand contrôle sur les détails du rendu apporterait des possibilités supplémentaires de sonification pour les environnements interactifs [Pau11], notamment pour les environnements urbains [FJ10].

2.4.1.4 Bilan

Les modèles informés et de synthèse environnementale offrent potentiellement la possibilité de reproduire n'importe quel type de source et seraient privilégiés au premier abord pour une application de type simulation. Cependant, la quantité de calculs et de traitements requis en amont rendent tant la production que le rendu difficiles à intégrer dans le cadre d'un média temps-réel interactif.

De plus, le rendu sonore le plus proche de la réalité n'est pas nécessairement celui recherché par le designer sonore, il convient parfois d'exagérer certains traits sonores pour accentuer la sensation recherchée (enveloppement, distance, vitesse, densité,...) ou de transformer le contenu fréquentiel des sons pour accentuer la lisibilité du mixage. Malgré ces limitations contextuelles, les modélisations physiques offrent des matériaux sonores crédibles et pouvant s'intégrer aux outils des

designers sonores [Far10], en parallèle de l'avancée des méthodes procédurales [Far07a], y compris pour la musique interactive [Col09].

L'ouvrage *Designing Sound* de Andy Farnell [Far10] offre un vaste panel des méthodes de synthèse disponibles ainsi que de leurs applications, en privilégiant de choisir une méthodologie adaptée à la texture recherchée, toutes les méthodes de synthèse n'étant pas adaptées à toutes les sources et matières.

Les méthodes de synthèse FM ou par tables d'ondes ne semblent pas non plus adaptées à la production efficace de tels rendus. En revanche, les méthodes de synthèse granulaire semblent proposer des stratégies offrant un compromis entre la ressource mémoire occupée par les échantillons sonores et le temps de calcul nécessaires à la granulation et aux divers traitements appliqués en temps-réel aux grains. Il convient cependant, lors du développement d'une application de synthèse granulaire, de porter une attention particulière à la méthode de granulation choisie au regard des fichiers sonores à traiter.

Les méthodes de synthèse granulaire, dans le domaine temporel, semblent apporter un compromis intéressant entre la quantité de calcul requise et les perspectives de résultats.

Dans les paragraphes suivants, nous nous intéressons aux méthodes de synthèse sonore présentes dans les outils dédiés à la sonorisation des environnements virtuels de type jeu vidéo. Ces méthodes nous intéressent particulièrement dans la mesure où, dans les jeux vidéo, les besoins de sonorisation sont aussi importants qu'ils sont contraints en ressources allouées. Notre travail de thèse portant sur le développement d'un outil de synthèse sonore temps-réel dédiée aux villes virtuelles, il nous paraît essentiel de mettre en regard les méthodes de synthèse complexes (que nous avons brièvement décrites) avec les outils de l'industrie vidéoludique. Cette mise en regard a pour but de dégager un compromis pour la stratégie de conception de notre outil de synthèse.

2.4.2 Outils dédiés au jeu vidéo

Par certains aspects, la production sonore d'un jeu vidéo ressemble à celle d'un film [Col08]. De nombreuses techniques d'enregistrements sont communes aux deux médias (bruitages, voix), il en va de même pour le choix des logiciels d'enregistrement et d'édition audio et les lignes directrices de la composition des accompagnements musicaux. Il existe cependant des différences très nettes. Si le son d'un film repose

beaucoup sur l'étape de post-production, la nature interactive du jeu vidéo ne permet pas la mise en place du même processus de création. De même, les ambiances sonores utilisées pour un film sont souvent enregistrées sur le lieu de production, en parallèle du tournage. Cette manière de procéder ne peut être appliquée à la production d'un jeu, il faut alors reconstituer les ambiances en studio (ce compromis n'est pas toujours accepté dans la production cinématographique, il est parfois suspecté de « falsifier l'espace ») ou à l'aide de *field recordings*.

L'étape de mixage est également plus complexe dans les jeux vidéo que dans le cinéma car elle doit tenir compte des changements en temps réel liés au *gameplay* tout en préservant les qualités d'immersion de l'environnement proposé [Tho10]. Le mixage d'un jeu vise à anticiper ces changements. De fait, la bande son d'un jeu, y compris sa musique, n'est pas d'un seul tenant, elle est constituée de fragments sonores sélectionnés, lus et mixés en temps réel tout au long de l'expérience de jeu. Deux exemples de ces différences sont donnés dans [NLP13] :

Considérons, par exemple, un son attaché à une source visuelle fixe, comme le son d'une fontaine. Dans un film, le réalisateur décide avec le monteur à partir de quel plan et de quel time code le son de la fontaine doit être perçu. La variation de son volume est mixée de façon déterministe en suivant le déplacement de la caméra. Dans un jeu, on ne connaît à l'avance ni la position de la caméra, ni le point d'écoute. Le son de la fontaine est positionné sur la carte du jeu à un endroit précis, indépendant de la position de l'avatar. Ce son est doté d'un cône d'atténuation déterminant à partir de quelle distance le joueur commence à entendre la fontaine et quelle est la variation de l'intensité perçue en fonction de la distance entre l'avatar et cette fontaine.

Prenons, comme second exemple, un son attaché à un objet mobile, comme le son des pas de l'avatar du joueur lorsque l'objectif qui lui est assigné est de pénétrer une forteresse. Dans un film, suivant le scénario et le montage, ce son sera choisi et mixé en suivant une évolution contrôlée et subtilement calculée. Dans un jeu, le scénario est loin d'être complètement écrit : le joueur contrôle le déplacement de l'avatar, donc la caméra, et peut ainsi choisir différentes stratégies d'attaque. Le désigner crée un flux sonore de pas, plus ou moins filtré en fonction de l'action de l'avatar, tex-

turé selon la nature du sol sur lequel il se déplace et adapté au type et à la vitesse du mouvement (sur place, marche, course, saut...). De plus, ce son est lié au contexte fictionnel. Si le joueur opte pour une stratégie d'infiltration, il faut générer des bruits de pas discret. A contrario, s'il décide d'utiliser la méthode brutale et de tirer sur tout ce qui bouge, on accentuera les sons de déplacement.

Le jeu se déroulant en temps réel, d'autres enjeux techniques liés aux ressources informatiques influencent la composition de la bande-son. De manière générale, deux à cinq pour cent du temps de calcul processeur et jusqu'à dix pour cent des ressources mémoire [Bra08] sont allouées aux aspect sonores d'un jeu. Ce ratio dépend aussi du contenu du jeu : la production d'un jeu sonore attribuera plus de ressources à l'audio qu'un jeu d'action classique.

Des outils de production sonore sont donc spécialement conçus pour répondre aux enjeux et contraintes de la production sonore des jeux vidéo. Dans les paragraphes suivants, nous présentons les principales solutions audio disponibles ainsi que des techniques procédurales émergentes.

2.4.2.1 Fmod et Wwise

Les deux principaux moteurs sonores de jeu vidéo du marché sont actuellement Fmod (produit par Firelight Technologies¹²) et Wwise (produit par Audiokinetic¹³). Dans leur majorité, les moteurs utilisés pour le développement de jeux vidéo (Unreal Engine, CryEngine,...) proposent des solutions audio intégrées. Cependant, Fmod et Wwise sont des outils indépendants très avancés (notamment en comparaison d'autres solutions comme OpenAL) en raison de leurs composants spécifiques. Notons également l'utilisation d'une partie des technologies de Fmod dans le moteur Unity3D¹⁴.

Fmod et Wwise proposent une interface graphique conçue dans le but de faciliter le travail des designers sonores. L'interface graphique donne accès à des paramètres permettant de façonner le comportement des événements sonores en fonction de variables entrantes. Dans cette interface, le designer sonore choisit quels sons seront joués, de quelle

12. <http://www.fmod.org/>

13. <https://www.audiokinetic.com/products/208-wwise/>

14. <http://unity3d.com/>

façon ils le seront (une seule fois ou en boucle, en séries linéaires ou bien aléatoires) et quels traitements leur sont éventuellement appliqués (filtres, spatialisation et autres effets).

Une deuxième interface, plus bas niveau, est celle qui, par le code, relie le moteur sonore au reste de l'environnement virtuel. Ce lien fait transiter les variables renseignant sur le déclenchement des événements sonores, la position des sources sonores dans l'espace du jeu et tout autre paramètre interfacé. Toutes ces informations servent donc autant au déclenchement des fichiers audio qu'au mixage temps-réel de la bande-son qu'ils composent. De nombreuses itérations sont parfois nécessaires afin d'obtenir la calibration souhaitée des paramètres sur les transformations sonores [Col08].

En plus des capacités de lecture de fichiers audio précédemment décrites, *Fmod* et *Wwise* proposent certaines méthodes de synthèse sonore, essentiellement de la synthèse soustractive et certaines fonctionnalités proches de la synthèse granulaire.

La synthèse soustractive consiste à soustraire, à l'aide de filtres, une partie d'un signal, le plus souvent généré par des oscillateurs ou des sources de bruit. Elle est principalement utilisée pour la création d'effets « *woosh* » ou d'autres sons percussifs simples, le plus souvent à partir d'un bruit blanc passé au travers d'un filtre résonnant. Ces méthodes basiques ne sont pas suffisantes pour concevoir à elles seules des paysages sonores complexes tels que ceux d'une ville.

Des méthodes de synthèse granulaire simples sont également implémentées dans *Fmod* et *Wwise*. Leur utilisation permet principalement d'apporter des variations et de l'aléatoire sur des sons d'impact, des bruits de pas, des frottements ou certaines pistes d'ambiances (comme des chants d'oiseaux ou des fonds d'air). Les outils de synthèse granulaire disponibles dans *Fmod* et *Wwise* sont économes en temps de calcul mais semblent encore limités du point de vue de la sophistication du rendu qu'ils proposent ainsi que dans le choix des paramètres de granulation.

Des deux types de synthèse présentes dans les moteurs sonores *Fmod* et *Wwise*, la synthèse granulaire semble la plus pertinente pour satisfaire aux besoins complexes de la sonorisation d'une ville virtuelle. Cependant, la forme de granulation disponible dans ces moteurs n'est pas spécialement dédiée aux besoins d'une synthèse environnementale riche et complexe. La thèse d'Olivier Veneri [Ven09] propose un état

de l'art très complet sur les principales caractéristiques des moteurs sonores de jeu ainsi que sur leur communication avec l'environnement du jeu. Nous notons cependant ici qu'aucun outil n'est spécifiquement dédié à la création des ambiances d'arrière-plan. Comme nous l'avons vu dans la section précédente, leur construction se fait principalement *via* des stratégies basées sur la lecture aléatoires de fichiers audio ou la constitution de boucles mixées dynamiquement. Dans les paragraphes suivants, nous nous intéressons donc à la question des stratégies dites *procédurales* dans la sonorisation des jeux vidéo.

2.4.2.2 Audio procédural et jeux vidéo

Comme nous l'avons vu dans les paragraphes précédents, Fmod et Wwise proposent des outils pour appréhender ponctuellement de la synthèse en temps réel. Il s'agit principalement d'oscillateurs virtuels ou de sources bruitées et de séries de filtres dédiés à de la synthèse soustractive. Ces moteurs permettent aussi d'appliquer des principes de synthèse granulaire à certains échantillons. Il s'agit ici principalement de contourner de potentiels effets et de répétition nuisibles à la sensation d'immersion du joueur. L'idée d'un design sonore intégralement procédural reste marginale dans la production du jeu vidéo.

Cependant, ainsi que le présente Andy Farnell [Far07a], l'outil procédural possède des capacités qui vont au-delà de ce qu'il est possible de réaliser avec des fichiers sonores linéaires :

Procedural audio is non-linear, often synthetic sound, created in real time according to a set of programmatic rules and live input. To further define it let's explain it in terms of linear, recorded, interactive, adaptive, sequenced, synthetic, generative and AI audio. Let's also analyse some other phrases like algorithmic composition and stochastic music. By doing so, we may reach an understanding of why procedural audio is so powerful and important.

L'utilisation de méthodes procédurales dans les jeux vidéo, appliquées aux animations ou aux *assets* graphiques, s'est accrue au cours des dernières années. Cette démarche est encore faiblement utilisée pour le son [Col08].

La nature interactive d'un jeu vidéo et le nombre important de paramètres régissant son univers virtuel constituent autant de points d'entrée pour faire varier, en temps réel, la synthèse. Les méthodes

procédurales peuvent mettre à profit cette interaction constante entre l'univers du jeu et son moteur sonore pour créer des bande-sons finement réactives et adaptatives. Le jeu *Spore* (Electronic Arts, 2008) constitue un des exemples de bande-son procédurale dans laquelle les facteurs de synthèse permettent de créer une grande variété de matières sonores. Le type de synthèse utilisée est choisi selon les résultats sonores recherchés. Si potentiellement plusieurs méthodes de synthèse sonore permettent d'obtenir un même rendu, le coût de calcul sera différent pour chacune. Il convient donc de choisir la méthode appropriée à chaque type de son [Far10]. La synthèse modale [BDT⁺08], la synthèse granulaire [Pau11] ou des modèles informés [VAKMP10] ont, entre autres, fait l'objet d'études sur leur application à des environnements interactifs.

En parallèle, la musique dite *générative*¹⁵ fait l'objet de recherches appliquées aux jeux vidéo [Col09]. Il s'agit de concevoir des procédés temps-réel de composition musicale réagissant aux paramètres du jeu [Ven09]. L'objectif de ces techniques est principalement d'enrichir l'expérience du joueur en concevant des environnements sonores proposant plus de réactivité aux interactions et moins de risques de répétitions dans les événements entendus.

La réalisation audio d'un jeu vidéo est un travail collaboratif entre plusieurs disciplines. Chaque spécialité (design sonore, dialogues, ambiances,...) doit prendre en compte les autres et l'ensemble de la bande-son n'existe efficacement que dans son interaction avec les autres entités du jeu (graphisme, code,...).

Les mécanismes et stratégies d'intégration des éléments sonores sont aussi importants que la composition des éléments elle-même car ils régissent la manière dont l'ensemble de la bande-son sera perçue.

La répartition des ressources allouées aux échantillons ou aux traitements temps réel est la source d'un débat ouvert dans la production d'un jeu [Fou10]. Les échantillons puisent dans les ressources mémoire tandis que les traitements temps réel mobilisent du temps de calcul. Si les techniques de synthèse en temps réel ne peuvent aujourd'hui pas remplacer toutes les sources sonores d'un jeu ou la couleur sonore du travail d'un designer, elle peuvent être ponctuellement utilisées et ap-

15. A ce titre, les jeux « sonores » [Gau08] font l'objet de développements particuliers [FG04].

porter des variations qu'il ne serait pas possible d'obtenir par le seul mélange d'échantillons.

Nous avons vu que la détermination de la nature de l'objet sonore à produire est un facteur essentiel dans la réalisation de celui-ci. Ses propriétés de timbre, l'anticipation de son évolution dans le temps et ses interactions avec l'environnement sont autant de facteurs à prendre en compte dans la stratégie de réalisation. Il nous paraît donc essentiel, dans la section suivante, de délimiter plus nettement les contours de l'objet sonore arrière-plan dans le but de concevoir une stratégie de sonorisation temps-réel efficace, offrant un rendu varié et interactif.

2.5 L'arrière-plan urbain comme texture sonore

Dans les sections précédentes, la notion d'arrière-plan sonore est traitée au travers du prisme de plusieurs disciplines et méthodes de sonorisation. Il en ressort que l'arrière-plan peut être compris comme un objet sonore à part entière et constitue une part non négligeable dans le discours d'une scène sonore urbaine interactive.

Nous souhaiterions maintenant effectuer un rapprochement entre l'arrière-plan, en tant qu'objet sonore, et des méthodologies de (re)synthèse propices à sa génération. La majorité des méthodes de sonorisation que nous avons étudiée dans les sections précédentes montrent que les arrière-plans sont souvent sonorisés par des fichiers pré-enregistrés mis en boucle. Nous émettons cependant l'hypothèse que les arrière-plans sonores urbains peuvent être sonorisés à l'aide de méthodes procédurales plus riches dans leur rendu dynamique.

Nous proposons donc ici de nous intéresser plus particulièrement à la notion de texture sonore ainsi qu'aux différentes méthodes de synthèse associées. Nous présentons, dans un premier temps, les définitions clefs relatives au concept de texture sonore puis passons en revue différentes méthodes de (re)synthèse associées et, enfin, une conclusion mettant en rapport texture et arrière-plans sonores.

2.5.1 Définition d'une texture sonore

Les textures sonores sont des objets importants dans la sonorisation de jeux vidéo, de films, de simulations, d'installations interactives, de réalités virtuelles ou d'applications web, notamment pour la production de sons environnementaux [Gri10]. L'enjeu de leur synthèse est de mettre en place une stratégie qui permette de générer des matériaux sonores d'une durée supérieure et de timbre plus varié que le matériau de référence pré-enregistré [SER06].

Il n'existe pas de définition fermée de ce qu'est une texture sonore mais davantage une série de propriétés. Une des premières proposition d'analyse et de synthèse des textures sonores fut formulée par Saint-Arnaud et Popat en 1998 [SAP97]. Ils définissent une texture sonore présentant des caractéristiques stables dans le temps. Elle peut contenir des structures ponctuelles ou des éléments d'aléatoire, mais

la structure, prise dans son ensemble, doit rester stable. Une texture sonore est caractérisée par son *sustain*, son maintien dans le temps.

Cette caractérisation dynamique est représentée sous forme d'un schéma (Fig. 2.18) qui présente des caractéristiques analogues à celui communément utilisé pour décrire le drone urbain (voir plus loin Fig. 3.2).

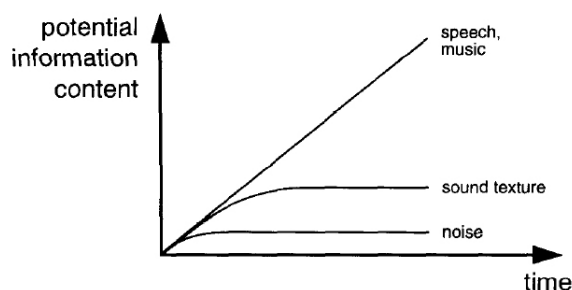


FIGURE 2.18 – Le bruit et les textures sonores montrant des caractéristiques constantes sur le long-terme, *in* [SAP97]

Saint-Arnaud et Popat mettent également en avant qu'une texture sonore ne comporte pas de brusques changements de tonalité et le rythme qui la sous-tend ne doit ni accélérer ni décélérer. Les sons qui sont fortement caractérisés par des éléments transitoires ne peuvent constituer des textures. Ils proposent de comprendre une texture sonore comme un ensemble bas-niveau d'atomes distribués et organisés selon une structure haut-niveau. Les caractéristiques haut-niveau d'une texture doivent être pleinement perçues dans une fenêtre d'attention de quelques secondes. Cette dernière affirmation nous intéresse particulièrement dans la mesure où elle pose la question de la frontière entre un son individuel (une voiture) et une texture (le trafic) [SER06], rejoint la problématique de la définition perceptive des frontières d'un arrière-plan ainsi que de son articulation avec les sons d'avant-plan.

Strobl et al. [SER06] proposent également une classification sémantique des textures sonores. Ils distinguent les textures naturelles (feu, eau, vent), des textures animales (goélands, grillons), les textures humaines (bavardages), les textures mécaniques (drones, trafic, martellements, bourdonnements, grondements), les textures exprimant une activité (pas, touches d'ordinateurs, frottements, claquements, grattements, glissements, crépitements). D'autres approches perceptives sta-

tistiques sont développées afin de classer les textures [EZM11] ou d'en comprendre les mécanismes de perception [MS11].

Il existe donc une variété de textures sonores [Gri10] et de méthodes pour les (re-)synthétiser. Nous présentons ces dernières dans les paragraphes suivants.

2.5.2 Synthèse de textures

De même qu'il n'existe pas une définition fermée de la notion de texture sonore, il n'existe pas une méthodologie unique pour leur synthèse. Thomas Grill [Gri10] regroupe les méthodes de synthèse de textures sonores en trois grandes catégories :

- Les méthodes basées sur la décomposition multi-résolution du signal, utilisant principalement les transformées en ondelettes.
- Les méthodes basées sur la décomposition temporelle du signal, recouvrant notamment la synthèse granulaire.
- Les méthodes reposant sur des représentations source-filtres d'un signal modèle.

Ces méthodes sont principalement employées avec pour but la production d'une texture convaincante sur la base d'échantillons pré-existants.

Les méthodes d'analyse/re-synthèse mises en œuvre pour reproduire les textures partent donc le plus souvent d'un échantillon pré-enregistré et cherchent à répondre aux trois questions suivantes :

- Quelle stratégie de segmentation adopter sur l'échantillon ?
- Quels paramètres extraire des segments sélectionnés ?
- Quelle est la meilleure méthode de synthèse à adopter avec pour objectif de créer un nouvel échantillon qui soit à la fois d'une durée plus longue que l'original mais d'une qualité équivalente ?

Il n'existe à l'heure actuelle que peu de moteurs de synthèse dédiés aux textures sonores [Sch11] et la plupart d'entre eux sont spécifiques à certaines textures. Cependant, le principe proposé par Saint-Arnaud et Popat dans [SAP97] reste le plus utilisé. Il consiste à distinguer deux niveaux de resynthèse : la partie bas-niveau a pour objectif de créer des atomes sonores à partir d'un fichier source pré-enregistré et la partie haut-niveau est relative à la distribution et à l'organisation de ces atomes.

Diemo Schwarz distingue la synthèse de *textures expressives* et la synthèse de *textures naturelles* [SS10]. Le premier type vise l'interactivité et la génération de sons pour des compositions ou des performances

musicales, le plus souvent sous la forme d'un instrument digital. Le second type vise à reproduire des paysages sonores plus larges dont le principal critère est le réalisme, la crédibilité des textures obtenues repose en grande partie sur les ambiances et les informations restituées.

Hoskinson et Pai, dans [HP01], proposent en 2001 une des premières applications de synthèse avancée de textures sonores. Il s'agit de segmenter un enregistrement vocal pour recréer une version aléatoire du signal, dans un flux continu, tout en préservant un maximum des caractéristiques d'origine de l'enregistrement. L'échelle bas-niveau (soit la taille des grains) choisie pour la re-synthèse est celle de la syllabe, principalement afin de rendre l'analyse et la transition entre les grains plus aisées. L'enchaînement entre les grains, soit la partie haut-niveau de la synthèse, repose sur une analyse de similarité afin de trouver des points de transition les plus naturels possible.

En 2007, Strobl [Str07] propose un générateur de textures sonores, avec analyse et re-synthèse, réalisé dans Pure Data. Strobl reprend la définition du flux audio proposée par Bregman [Bre94] : une unité perceptive d'un événement, ce qui amène un regroupement des composantes individuelles du flux (comme une série de bruits de pas par exemple). Strobl décrit de façon plus générale les propriétés d'un « flux de texture sonore naturelle ». La texture continue obtenue par réorganisation des grains contient les même flux audio que le son d'origine, ce dernier est simplement arrangé dans un ordre différent. Dans la continuité de Saint-Arnaud et Popat [SAP97], l'adjectif « naturel » employé ici par Strobl suggère que le flux audio de re-synthèse n'est pas une simple juxtaposition des fragments sonore mais une séquence lisse de sons perceptivement significatifs les uns par rapport aux autres.

Dans une analyse récente et particulièrement complète, Diemo Schwarz fait état de deux modèles de synthèse [Sch11] dédiés aux textures sonores fonctionnant particulièrement bien. La première méthode proposée par Diemo Schwarz et Norbert Schnell dans [SS10] est basée sur la synthèse concaténative par corpus. Elle consiste à analyser un fragment d'un enregistrement à l'aide des descripteurs dans CataRT. Les descripteurs fournissent une définition statistique de référence pour le positionnement dans un espace virtuel à deux dimensions des segments à resynthétiser. Le corpus ainsi constitué est ensuite étendu, avec des versions transposées et filtrées du matériau de départ, afin de couvrir au maximum les variations des descripteurs. La seconde méthode, pro-

posée par O'Regan et Koharam dans [OK07] est une adaptation à une dimension de l'algorithme de Efros et Lung [EL99] dédié à la synthèse de textures d'images non-paramétriques en deux dimensions.

Comme nous l'avons vu dans les paragraphes précédents, les méthodes de (re-)synthèse de textures sonores sont dépendantes du matériau de départ et de son contexte d'utilisation. Qu'en est-il alors des textures sonores propres au milieu urbain ?

Finney [FJ10] propose un générateur de textures dédiées aux ambiances urbaines et conçu pour *Google Street View*¹⁶. Finney, reprenant la segmentation de Murray Schafer, distingue la texture d'arrière-plan et les objets de premiers plans. Les sons de base servant pour le modèle sont extraits de la banque de données *Freesound*¹⁷. Les textures et les objets sonores sont extraits et répartis entre les deux catégories sur la base de leurs identifiants sémantiques. Ces derniers proviennent soit des *tags* descriptifs fournis par *Freesound* soit d'un modèle d'analyse de type MFCC-BIC. Un premier tri a cependant été effectué dans la banque de *Freesound* en sélectionnant autant que possible des enregistrements de Barcelone, la ville qui a servi de support visuel pour les évaluations. Malgré la procédure d'analyse et de tri, les objets sonores sont ensuite récupérés manuellement. Au final, l'ensemble du paysage sonore urbain est synthétisé à l'aide de trois types de sources : des échantillons lus, de la re-synthèse par ondelettes et de la synthèse concaténative. Finney justifie ce choix en argumentant que combiner plusieurs méthodes permet une séparation optimale des sources et une plus grande efficacité dans le traitement de chaque type de sons.

Le procédé d'évaluation, qui a consisté à faire des écoutes comparatives entre trois paysages sonores générés et trois enregistrements, en quadriphonie, montrent une préférence pour le synthétiseur. Considérant que les questions d'évaluations portaient principalement sur l'identification, la localisation et le dynamisme des objets sonores, Finney pose l'hypothèse que cette préférence se justifie par la spatialisation et le mixage des deux strates sonores qui permettent plus facilement de distinguer et de localiser les objets dans la scène sonore qui est elle-

16. Au moment où nous rédigeons ce document, le lien Internet permettant d'accéder au démonstrateur de l'outil développé par Nathaniel Finney ne semble plus disponible.

17. <http://www.freesound.org/> *Freesound* est un projet de bases de données sonores, collaboratives et gratuites sous le régime des licences *Creative Commons*.

même, de fait, plus dynamique. En comparant la description faite par Finney et les principes qualitatifs édictés par l'écologie sonore (présentés dans le Chapitre 2), le paysage généré semble plus proche d'une définition *hi-fi* que *lo-fi*, cette dernière appellation étant généralement attribuée aux environnements urbains réels.

Le recours à la séparation des strates sonores, notamment pour la spatialisation et la gestion des niveaux sonores, semble donc être une technique efficace pour créer des environnements urbains virtuels immersifs. La ressemblance parfaite avec la réalité semble moins prise par l'auditeur que la clarté du mixage et la qualité perceptive de l'écoute.

2.5.3 Arrière-plans sonores : textures complexes

Dans son principe général, la synthèse de textures consiste à étendre les possibilités de lecture d'un son composé de micro-événements dont les caractéristiques perceptives restent stables dans le temps. Le but de la re-synthèse un tel son est de pouvoir produire des variations sur une durée indéterminée, sans artefacts, tout en conservant les propriétés du fichier audio d'origine. Le concept de *sound textures* (textures sonores) auquel nous faisons référence ici diffère cependant de celui de *audio textures* (textures audio) [OK07]. La synthèse de textures audio [LWZ04] consiste à générer des séquences après analyse des fichiers, l'analyse déterminant par similitude les points de transitions possibles entre les fichiers. Parallèlement, la synthèse de textures sonores consiste à générer une textures nouvelle et plus longue sur la base des fichiers originaux.

Les propriétés morphologiques des arrière-plans sonores urbains (continuité, stabilité dynamique, ubiquité) que nous avons étudiées dans les sections précédentes semblent permettre de les envisager comme des textures sonores pour lesquelles le développement d'un outil de (re-)synthèse reste à définir. Les méthodes de synthèse de textures sonores présentées offrent un panel varié de stratégies envisageables. Il ne faut cependant pas perdre de vue le contexte interactif et temps-réel d'une telle réalisation, ce qui pose des contraintes en termes d'utilisations de ressources. Certaines méthodes d'analyses coûteuses en temps de calcul sont donc exclues. En revanche, le principe de la synthèse de texture (créer, à partir d'un échantillon, des variations plus longues que l'échantillon lui-même) est une perspective d'économies de ressources

mémoire et de plus grandes possibilités d'interactions, relativement à un fichier d'ambiance mis en lecture bouclée.

Le choix de la méthode de synthèse à appliquer pour les objets sonores de type arrière-plan est relatif à la manière de caractériser. Les propriétés temporelles, timbrales et morphologiques présentées dans les chapitres précédents nous amènent logiquement à envisager le drone urbain comme un objet sonore complexe pouvant posséder plusieurs strates de textures. Pour Farnell, les arrière-plans d'une scène sont composés de textures [Far07c] et ces textures sont faites de sons possédant des éléments structuraux qui se répètent dans le temps avec de l'aléatoire.

2.6 Conclusion

Au travers de ce chapitre, nous avons pu mettre en évidence la singularité de l'arrière-plan urbain en tant qu'objet sonore. L'arrière-plan apparaît comme une composante à part entière de l'environnement sonore urbain. Sa description dans la théorie du Paysage Sonore est reprise pour la conception de scènes pour le cinéma ou les environnements virtuels comme les jeux vidéo.

L'arrière-plan possède un certain nombre de propriétés perceptives, morphologiques (diffusion acoustique, spectre contenant peu de fréquences aiguës, stabilité dynamique, variations sur une échelle temporelle relativement lente) et tient une place à part dans la narration d'une scène sonore. Ses propriétés lui confèrent des processus d'élaboration et d'intégration qui lui sont propres. La dynamique de l'arrière-plan est souvent plus lente que celle des objets individuels. Cependant, il s'agit d'un objet variable et adaptable, tant dans les médias linéaires que non-linéaires.

Le recours à des plans sonores différenciés est courant au cinéma (notamment par l'enregistrement d'ambiances servant de support aux dialogues et autres bruitages d'une scène), cette stratégie se retrouve fréquemment dans les jeux vidéo et les environnements virtuels. Les ambiances et arrière-plans sont un support essentiel sans lequel le rendu visuel peut apparaître terne et artificiel [JBM13]. Nous posons l'hypothèse que la segmentation d'une scène sonore en plans successifs peut servir de base à des stratégies de sonification en créant plusieurs niveaux d'écriture sonore et d'interaction dans le rendu de la scène. Le designer sonore peut utiliser ce niveau d'écriture supplémentaire dans la construction de la scène pour dispenser des informations qui ne peuvent être traduites par la somme des sources individuelles. Cependant, la sonification des arrière-plans repose encore majoritairement sur l'utilisation de larges échantillons mis en boucle. Des méthodes procédurales plus avancées permettraient d'apporter de la diversité dans le rendu et d'accentuer la flexibilité et la réactivité de la génération des ambiances.

L'arrière-plan doit, le plus souvent, occuper un nombre suffisant de canaux dans le rendu pour englober les autres sons de la scène (par exemple, les arrière-plans du jeu Prototype sont conçus en quadriphonie dans le cadre d'un rendu au format 5.1). L'ubiquité permet de

créer l'espace de perception de l'arrière-plan, celui-ci devenant une empreinte de l'espace de la scène. Les informations dont il est le support passent donc essentiellement par des variations dans les textures qui le composent. Dès lors, il convient de déterminer comment l'arrière-plan interagit avec l'environnement virtuel ainsi qu'avec les déplacements de l'écoutant dans cet environnement. Un arrière-plan sonore est une source audio avec des propriétés singulières. Il caractérise l'ensemble de la scène, son état, son atmosphère, relativement à la position du point d'écoute. Les deux principales questions auxquelles le designer sonore doit répondre pour concevoir les mécanismes dynamiques d'un arrière-plan sont donc : Quels sont les descripteurs de la scène virtuelle nécessaire à l'évolution de l'arrière-plan, relativement aux informations transmises par ce dernier ? Comment les déplacements du point d'écoute modifient la perception de l'arrière-plan ?

Dans une ville virtuelle, la gestion de centaines de sources dynamiques (piétons, véhicules,...) pose des problèmes d'optimisation (polyphonie maximale supportée par le moteur audio, calculs de spatialisation de chaque source,...). Des stratégies de groupement dynamique de sources peuvent constituer des alternatives mais elles restent coûteuses en temps de calcul [MBT⁺07]. Il serait aussi possible de s'interroger sur la nécessité d'accorder la même quantité de ressources processeur à une source proche du point d'écoute et à une source lointaine qui, de plus, a de fortes chances d'être masquée. Il ne s'agit cependant pas d'ignorer purement et simplement les sources lointaines. L'utilisation d'un arrière-plan consiste à substituer les sources individuelles les moins pertinentes (donc celles définies comme trop lointaines pour être perçues distinctement) par un flux audio unique. La question de la ligne de séparation géographique entre les sources individuelles et l'arrière-plan est centrale dans la stratégie de sonification : il faut définir quels matériaux sonores sont intégrés à l'arrière-plan et quels sont les mécanismes, dans le mixage dynamique de la scène, qui lient l'arrière-plan aux autres sources. Notre hypothèse est que la réponse à cette question est propre à chaque environnement virtuel, en fonction des paramètres d'interactivité et des autres composantes de l'environnement (intelligence artificielle, scénario,...). Un outil polyvalent de création d'arrière-plans sonores devrait donc être conçu en prévision de multiples usages.

Nous avons également pu observer les liens entre l'arrière-plan en tant qu'objet sonore et la notion de texture sonore. Ces liens sous-

entendent la possibilité de développer des méthodes de synthèse sonore dédiées aux arrière-plans urbains. Pourtant, de telles méthodes restent peu explorées et la production vidéoludique privilégie l'utilisation de larges échantillons par rapport à des méthodes procédurales potentiellement plus économes en ressources.

L'essor des méthodes de synthèse appliquées aux médias interactifs laisse entrevoir la possibilité de développer des paradigmes sur-mesure pour la sonorisation d'environnements complexes tels que des villes virtuelles. Nous proposons donc d'utiliser les techniques de synthèse sonore afin de créer un outil dédié aux arrière-plans urbains.

Chapitre 3

Problématiques de recherche

Sommaire

3.1	Introduction	100
3.2	Trois problématiques	101
3.2.1	Quelles articulations pour les plans sonores ?	101
3.2.2	Quelles méthodes de synthèse sonore pour caractériser une ville ?	102
3.2.3	Quelles méthodes de rendu sonore sémantique ?	103
3.3	Points de vue sur l'arrière-plan sonore . . .	105
3.3.1	Le point de vue acoustique	105
3.3.2	Le point de vue perceptif	112
3.3.3	Le point de vue de la composition et de la mise en scène	114
3.3.4	Dynamique de l'arrière-plan	117
3.4	Conclusion	121
3.4.1	Caractéristiques de l'arrière-plan sonore urbain	121
3.4.2	Enregistrement de l'arrière-plan sonore urbain	122

3.1 Introduction

Ce chapitre présente les problématiques de recherche traitées au cours de la thèse, les enjeux qui y sont liés et les solutions envisagées. Dans le Chapitre 2, nous avons mis en avant de nombreux axes de recherche possibles concernant la synthèse sonore d’environnements urbains. Le Chapitre 3 a donc pour objectif de déterminer et d’argumenter l’axe de recherche suivi par la thèse.

L’objet sonore arrière-plan nécessite maintenant une recherche plus approfondie sur sa place relativement aux autres objets de l’environnement sonore urbain. La question fondamentale concerne ici la place de la texture d’arrière-plan par rapport aux autres objets sonores dans l’écriture d’une scène virtuelle. Implicitement, il s’agit également de définir les paramètres qui posent les frontières géographiques de l’arrière-plan à l’intérieur d’une scène.

Si nous avons jusqu’à présent considéré l’objet sonore arrière-plan comme une entité simple, il nous paraît important de poser la question de la complexité intrinsèque de cet objet, c’est-à-dire de la granularité de la texture d’arrière-plan relativement à son utilisation dans un environnement virtuel interactif.

Ces réflexions nous paraissent préalables au choix d’une méthode de synthèse, afin de s’assurer que celle-ci puisse répondre aux différents enjeux posés par la réalisation d’une texture sonore interactive. D’autre part, ce choix sous-entend de définir une stratégie dans la réalisation de textures. Notre préférence se portant vraisemblablement vers une méthode par granulation, la captation de fichiers sonores servant à alimenter le synthétiseur est elle-même soumise à certaines contraintes. Il convient donc de définir ces contraintes et de parvenir à une méthodologie d’enregistrement qui soit à même de produire des fichiers exploitables par le synthétiseur.

Sont décrites dans un premiers temps les trois problématiques centrales qui émergent de l’analyse des environnements urbains, dans la perspective de leur reproduction sonore. Après les avoir passées en revues et situé l’axe de recherche, nous présentons une analyse plus approfondie des arrière-plans sonores urbains avant d’exposer une proposition de méthodologie pour les reproduire et les intégrer à un environnement virtuel.

3.2 Trois problématiques

Le principal objectif de ce travail de thèse est de parvenir à la conceptualisation d'une solution pour la génération d'arrière-plans sonores urbains en temps réel. Le chapitre précédent éclaire les principaux concepts nécessaires à cette réflexion et décrit certaines solutions proposées précédemment. Le contexte des environnements virtuels tels que les jeux vidéo pose certaines contraintes de ressources et de réactivité auxquelles les méthodes procédurales semblent pouvoir bien répondre. Des alternatives aux boucles pré-enregistrées restent donc à explorer pour la gestion des ambiances urbaines.

La pluralité de cette thématique laisse entrevoir plusieurs axes de recherche et donc de problématiques. La théorie du Paysage Sonore fournit une base structurelle à la gestion d'ambiances sonores urbaines, les articulations entre les différentes strates sonores qu'elle décrit (comment les sons d'un environnement passent d'un plan à un autre ou d'une catégorie à une autre) constitue un axe de recherche en soi. Dans la continuité des travaux de recherche en audio procédural pour les jeux vidéo, l'utilisation des méthodes de synthèse pour la gestion des arrière-plans sonores urbains est un second axe. Enfin, il est également possible d'aborder la gestion des ambiances urbaines par l'axe des méthodes de rendu complexe telles que le groupement dynamique des sources.

Dans les paragraphes suivants, nous détaillons ces trois problématiques et argumentons celle que nous choisissons de résoudre, en lien avec les deux autres.

3.2.1 Quelles articulations pour les plans sonores ?

Un premier problème consiste à caractériser de façon opérationnelle les trois plans issus de la théorie du Paysage Sonore. Il s'agit de définir les principes perceptifs qui doivent être mis en œuvre dans un mécanisme temps réel gérant une scène urbaine afin de faire se déplacer les objets sonores d'un plan à l'autre.

Deux principaux types d'approches préexistent. La première, appelée *clustering de sources* [MBT⁺07], consiste à regrouper dynamiquement les sources sonores individuelles en groupes tout en anticipant les effets de masquage perceptifs. Cette méthode est implémentée dans le jeu vidéo de course automobile *Test Drive Unlimited* (Eden Games,

Atari, 2006) et permet de gérer plus d'une centaines de sources avec en moyenne une dizaines de *clusters*. Le seconde approche reprend la classification sémantique du Paysage Sonore [BMS05] en accordant à l'arrière-plan (*keynote*) une fonction définie de caractérisation de l'environnement virtuel.

Sur l'ensemble des méthodes existantes, le premier plan se distingue de l'arrière-plan. Une granularité plus fine dissocie les groupes des individus dans la scène virtuelle, ainsi que les catégories d'acteurs (humains, véhicules, sous-catégories telles que policiers, civils,...). Ce travail, réalisé depuis longtemps d'un point de vue visuel au cinéma et traité partiellement dans la synthèse graphique du jeu vidéo, reste à explorer d'un point de vue sonore dans les environnements virtuels. Cet aspect est à l'origine des deux autres problématiques détaillées ci-après.

3.2.2 Quelles méthodes de synthèse sonore pour caractériser une ville ?

Au Chapitre 2, nous avons examiné les méthodes de synthèse disponibles et montré que le problème de l'ambiance urbaine relève de la synthèse de textures sonores. D'autre part, nous nous plaçons dans un contexte d'applications de type jeu vidéo, ceci signifie que cette texture peut changer tant par son contenu que par la position du point d'écoute, de façon dynamique et imprévisible. Enfin, nous avons brièvement évoqué les contraintes de temps de calcul et d'espace mémoire qu'imposent le recours à ce type d'applications.

Le choix d'une méthode de synthèse doit donc répondre à ces différentes contraintes. Mais, pour être efficace, la méthode déterminée aura pour but de synthétiser un objet précis : l'ambiance d'arrière-plan d'une ville. Il faut donc, au préalable, identifier l'univers sonore qui caractérise une ville, principalement par le biais de son arrière-plan. Par « univers sonore » nous entendons les composantes individuelles du son urbain (les sources et leurs caractéristiques), leur comportement dans le temps (au fil de la journée) et l'espace (ici, acoustique de la ville) ainsi que dans leurs interactions.

Une fois l'univers sonore urbain circonscrit, et en prévision du contexte d'utilisation, nous pourrions faire précisément le choix d'une méthode de synthèse dans ses détails.

3.2.3 Quelles méthodes de rendu sonore sémantique ?

La gestion du rendu sonore temps réel qui articule dynamiquement ces différents plans constitue un troisième axe de recherche potentiel. L'enjeu est initialement lié à la complexité des algorithmes pour la spatialisation. Dans le contexte d'un environnement tridimensionnel graphiquement riche, il est difficile de spatialiser finement des dizaines de sources séparées. Une des principales solutions étudiées depuis une dizaine d'années reprend la notion graphique de niveaux de détails et le principe de regroupement (*clustering*) des sources.

Créer des niveaux de détails permet théoriquement de gérer le rendu sonore de plusieurs milliers de sources [TGD04] sans avoir à calculer les flux individuels de celles-ci. Ce besoin de segmentation est particulièrement important dans les scènes urbaines : celles-ci sont extrêmement denses en sources sonores (il est possible de compter des milliers d'acteurs) ainsi que par leur répétition (l'essentiel des sources étant des piétons ou des véhicules).

Un problème mal résolu est une définition à la fois efficace en terme de performance et perceptivement satisfaisante des critères de regroupement. Comme nous l'avons abordé dans le paragraphe sur les villes virtuelles (Chapitre 2, paragraphe 3.2), plusieurs stratégies existent, allant de la segmentation en groupes au clustering de source avec resynthèse en temps réel.

Ces points de vue, basés sur une approche acoustique, ne tiennent pas forcément compte de l'importance sémantique des sources, du point de vue de la composition et de la situation « narrative » de l'écouter. En effet, certains acteurs pourraient être regroupés en fonction de leurs caractéristiques (actions en cours, appartenance à une sous-catégorie,...) ce qui demanderait d'accéder à leur description sémantiques et de les grouper dynamiquement. Il est donc possible de dégager un troisième problème : la recherche de méthodes de groupement sémantique de sources sonores associée à un *clustering* dynamique.

Dans le cadre de cette thèse, nous nous sommes concentrés sur la seconde problématique. Toutefois, il est nécessaire de traiter en partie la première, car nous devons poser une définition cohérente de l'arrière-plan sonore urbain ainsi qu'envisager son articulation avec les autres plans. La relation au troisième problème est ouverte par le sens, la localisation, et donc le contenu que le compositeur veut donner à un

objet sonore variant de façon imprévisible dans le temps en fonction des actions d'un interacteur. Cette thèse a donc pour objectif de définir, spécifier et prototyper un outil de synthèse de l'arrière-plan urbain utilisable dans des applications fortement interactives telles que le jeu vidéo.

Ce choix se situe à la frontière des techniques mises en œuvre dans la gestion du rendu audio en temps réel et les principes conceptuels et perceptifs du design sonore. Ce problème a une application immédiate dans le projet Terra Dynamica, projet dans lequel nous avons pu intégrer et tester le prototype d'un outil de génération d'arrière-plan sonore urbain.

3.3 Points de vue sur l'arrière-plan sonore

La structuration d'une scène visuelle ou sonore en trois plans n'est de toute évidence pas d'origine physique. Elle trouve sa justification du point de vue perceptif (localisation, focalisation, sémantique des sources) et, dans le cadre d'une scène composée, de celui de l'agencement et de la mise en scène. Il est cependant possible de définir certaines des caractéristiques de ce qui est communément considéré comme l'arrière-plan sonore.

Dans cette section, nous proposons de présenter l'objet sonore arrière-plan, en milieu urbain, selon trois points d'écoute : acoustique, perceptif et composition. Ces trois points de vue sont complétés par une proposition du rapport dynamique entre l'arrière-plan et les autres objets sonores composant une scène.

3.3.1 Le point de vue acoustique

En tant qu'objet sonore, l'arrière-plan semble avoir une existence avant tout subjective. Dans le Chapitre 2, les sections relatives à recréation sonore des villes au cinéma et dans les jeux vidéo font apparaître le besoin de la mise en place d'un arrière-plan. Cette construction artificielle crée une évocation culturelle et une ambiance qui caractérise un lieu ainsi que l'état émotionnel qui lui est associé à un moment donné.

Cette idée est développée par Morgan [Mor09] dans le cas de la ville de New York. Morgan explique que New York est caractérisée par un arrière-plan sonore constant qu'il définit comme un drone présent en continu et qui ressort le plus distinctement sur les enregistrements les plus silencieux :

Although many of Manhattan's neighborhoods and boroughs have distinct and unique ambiances, what we discovered after several days of recording is Manhattan has a constant drone that underscores everything in the city. You can hear it in the parks, the subways and the busy streets – it is like a resonant note that plays continuously in the background, 24/7. Some of our quieter recordings reveal this keynote drone (...).

D'après Morgan, la captation de ce champ sonore lointain requiert de se trouver loin des sources sonores directes (sur des points en hauteur, dans des parcs ou des ruelles) et donc de ne capter que la diffusion

des sources sur l'environnement. Cette approche empirique est une tentative de ne capturer que le champ acoustique diffus de l'environnement urbain (voir 2.2.1.1, Figure 2.8), ce qui pose l'hypothèse de considérer le champ acoustique diffus comme une composante importante de l'arrière-plan sonore urbain.

Du point de vue de l'acoustique, il serait tout à fait envisageable de considérer l'arrière-plan d'un environnement comme étant la somme des réflexions acoustiques de chaque source visible ou hors du champs de vision. Pour Picaut *et al.* [PGD12], le milieu urbain est par nature un milieu diffus dans lequel les ondes sonores parviennent à l'auditeur depuis toutes les directions en raison du nombre important des réflexions. Augoyard circonscrit cet effet avec le terme « ubiquité », favorisé par la géographie sonore (« *sonic geography* ») de la ville ([Lab11] p.xxii). La théorie décomposant la densité d'énergie totale d'un champ sonore en champ direct et champ diffus (Fig. 2.8) date des travaux de Sabine [PGD12] et présente le champ diffus comme la résultante des réflexions du son sur les parois de l'environnement. Le champ diffus est distribué de façon égale dans le local de diffusion et pourrait donc y être observé depuis n'importe quel point. La structure temporelle caractéristique des événements acoustiques parvenant à un auditeur est schématisée Fig. 3.1. Le champ diffus est la réverbération tardive composée de réflexions rendues indissociables par leur forte densité temporelle ([Dan01], p.29). Ainsi que suggéré par l'effet d'ubiquité, les propriétés perceptives de la réverbération tardive dépendent davantage des caractéristiques du milieu de diffusion que de la position du point d'écoute (dans [Dan01], p.29, on parle également d'« isotropie » pour qualifier le champ sonore diffus). D'autre part, la présence du champ réverbéré tend à dégrader la capacité à localiser les sources sonores [SC00]. Si l'intensité du champ diffus reste constante, la position du point d'écoute fait varier l'intensité perçue du champ direct [Boo04], ces effets sont schématisés dans la Fig.3.2. Ainsi, en étant suffisamment éloigné des sources, il serait théoriquement possible de ne capter que le champ diffus d'un milieu.

Toutefois, Picaut *et al.* expliquent que l'hypothèse du champ diffus ne peut plus être vérifiée dans un milieu de propagation hétérogène et dont « l'absorption acoustique devient importante et non-uniforme ». Et, si le concept de champ diffus sert souvent de base aux tentatives de modélisation du son en milieu urbain, il n'y est pas directement transposable.

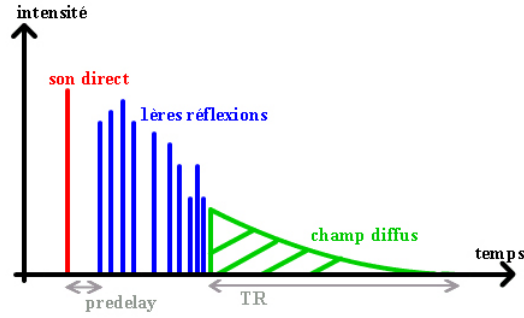


FIGURE 3.1 – Représentation de l'intensité dans le temps d'une réponse impulsionnelle

Notre approche perceptive de l'environnement urbain, dans ce travail de thèse, ne nous oriente pas vers une compréhension exhaustive des mécanismes acoustiques menant à la formation d'un fond diffus en milieu urbain. En revanche, il nous intéresse de tenter d'observer d'éventuelles constantes dans des enregistrements urbains. Ainsi que le suggère Morgan à propos de son travail sur le jeu vidéo *Prototype* [Mor09], ces constantes seraient susceptibles de fournir des pistes de réflexions pour définir une partie des éléments sonores caractérisant un lieu.

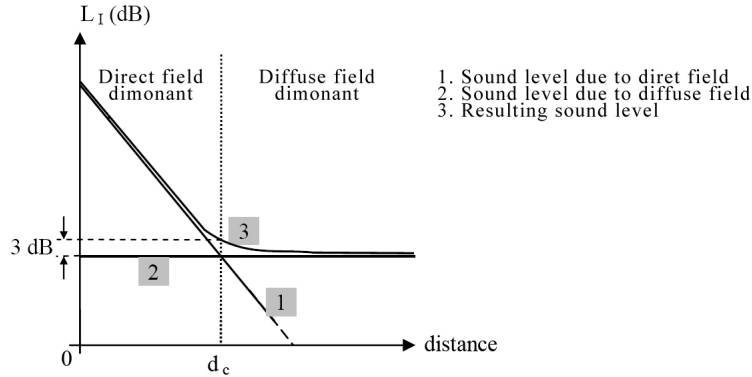


FIGURE 3.2 – Relations entre les niveaux sonores des champs direct et diffus

Notre travail de recherche nous amène donc à nous poser directement les deux questions suivantes : Peut-on parler d'un arrière-plan objectif, directement observable par l'analyse d'enregistrements ? Peut-

on donner quelques éléments qui donnent à croire qu'il existe objectivement une coloration spectrale de l'arrière-plan liée à un lieu? Le recours aux analyses spectrales est fréquent dans les champs d'étude du Paysage Sonore. Elles permettent dans un premier temps une représentation temps/fréquence simple afin d'analyser les événements constitutifs d'un enregistrement (voir section 2.1). Par ailleurs, certaines études visent aujourd'hui à automatiser l'identification des sources composant un enregistrement d'ambiance urbaine [DRRP06]. Dans les paragraphes suivants, nous présentons quatre séries d'enregistrements (initialement liés à la collecte de matières brutes et à l'écoute de différents aspects de l'environnement) accompagnées de sonagrammes (Annexe A). Chaque série se compose d'enregistrements effectués en différents endroits d'une même ville. Nous avons réalisé ces séries dans trois villes différentes Paris, Liège et San Francisco. Pour chaque série, les enregistrements sont effectués sur une journée, les conditions climatiques sont stables (pas ou très peu de vent, pas d'intempéries). L'objectif de ces analyses est d'observer, avec une approche naïve, d'éventuelles constantes sonores à l'intérieur de chaque série. Il n'est cependant pas question de comparer les séries entre elles.

3.3.1.1 Paris, de la station Quatre Septembre à Notre Dame de Lorette

Le premier enregistrement est une promenade sonore (ou *sound-walk*) réalisée entre les stations de métro Quatre Septembre et Notre Dame de Lorette. La représentation recouvre l'ensemble de l'enregistrement de huit minutes. L'itinéraire emprunté consistait essentiellement en une ligne droite, suivant un axe faiblement fréquenté (rue de Gramont, puis rue Lafitte) qui entrecoupe des axes plus fréquentés et plus larges (boulevard des Italiens, boulevard Haussmann, rue la Fayette, rue de Châteaudun).

La Fig.A.1 (Annexe A) est une représentation temps-fréquence de cet enregistrement, réalisée avec le logiciel Acousmographe¹. Y sont représentés le temps (en minutes, sur l'axe des abscisses) et une échelle de fréquence (en Hertz, sur l'axe des ordonnées). L'enregistrement que nous reportons ici ont été effectués avec un enregistreur stéréo portable Edirol R09HR.

1. <http://www.inagrm.com/accueil/outils/acousmographe>

En regardant cet enregistrement, on peut observer les variations de dynamiques induites par le passage des grands axes : une teinte noire plus prononcée indiquant une intensité sonore plus forte (à 2'30", 4'10" et 6'40" par exemple). On note également des trainées horizontales parallèles (1'10", 5'40") qui correspondent aux harmoniques de grincements de freins prononcés. D'autres événements sonores apparaissent aussi distinctement sur le spectre (klaxons, voix des passants, bruits de pas).

Il est possible de noter aussi certaines constantes sur l'ensemble de l'enregistrement. Une bande allant d'environ 20 à 3000 Hertz semble ressortir nettement. D'autres bandes de fréquences grisâtres apparaissent, notamment entre 4000 et 6000 Hertz, entre 11000 et 13000 Hertz. Ces bandes de fréquences sont d'une apparence quasi uniforme, évoquant davantage un bruit filtré qu'un son aux harmonies clairement discernables. Des accentuations apparaissent également à l'intérieur de ces bandes de fréquences, laissant entrevoir une structure sonore moins uniforme qu'un bruit blanc, plus complexe et colorée.

Alors que l'enregistrement se fait en déplacement, ces constantes restent présentes. Une écoute attentive du fichier ne semble pas révéler la présence d'artefact mais tendrait à laisser penser que ces constantes, ces trames, font partie du paysage sonore enregistré.

3.3.1.2 Belgique, Liège

Poursuivant dans cette direction d'écoute et d'étude, nous avons effectué des enregistrements et analyses dans d'autres situations et lieux. Les images A.2 à A.5 proviennent de différents enregistrements effectués dans la ville de Liège, en Belgique.

Les deux premiers sonagrammes (A.2 et A.3) proviennent d'enregistrements effectués sur les remparts de la ville, en se tournant vers deux quartiers différents de la ville depuis les hauteurs. Il s'agit de *field recordings* fixes. Les sonagrammes A.4 et A.5 sont des enregistrements effectués à deux endroits différents du centre ville. Sur les hauteurs des remparts, les sons urbains sont quasi exclusivement ceux captés de la ville à quelques centaines de mètres en contrebas. L'ascension vers les remparts ayant été faite à pied en suivant un petit chemin bordé d'arbres, les sons les plus proches sont donc principalement des sons de nature (des oiseaux discernables dans les aiguës, A.4). On note également des accentuations sur l'ensemble du spectre dans les figures A.4

et A.5, elles sont respectivement dû au passage d'un train (à quelques dizaines de mètres du lieu d'enregistrement) et à des travaux (dans le centre ville). On remarque la présence de traces harmoniques fréquentes dans les enregistrements lointains : il s'agit du son produit par les freins des bus de la ville, leur grincement portant sur plusieurs centaines de mètres.

La motivation des enregistrements effectués à Liège était de confronter les ambiances sonores de la ville perçue depuis des points éloignés à celles captées dans le centre ville lui-même. Malgré l'éloignement des différents points d'enregistrements les uns des autres (plusieurs centaines de mètres), les quatre enregistrements présentent une signature caractéristique dans les basses fréquences. Il s'agit d'un drone s'étendant des fréquences les plus graves jusqu'à environ 2500Hz. Le drone se divise en quatre accentuations régulières à l'intensité changeante selon les enregistrements : le drone semble devenir plus flou dans les enregistrements réalisés en centre ville. Ces derniers montrent des changements de dynamique et de timbre comparable à celui de la figure A.1 (Paris) tandis que le drone conserve sa régularité.

Nous observons une coloration du spectre des sonagrammes dans les basses fréquences, quel que soit le contexte environnant les enregistrements. Cette coloration prend la forme d'un drone qui varie légèrement selon la position du point d'enregistrement.

3.3.1.3 Paris, Bibliothèque Nationale

Nous avons réalisé une autre analyse comparative à Paris, cette fois à proximité de la Bibliothèque Nationale de France (BNF, site François Mitterrand), en confrontant deux enregistrements : l'un effectué sur le parvis de la bibliothèque, face à la Seine, (sonagramme A.6) et l'autre à quelques centaines de mètres de là, depuis la passerelle Simone de Beauvoir, au-dessus de la voie rapide du Quai de Bercy (sonagramme A.7).

Le contraste est volontairement accentué sur les deux rendus afin de faire ressortir les détails du bas du spectre. L'enregistrement effectué depuis le parvis présente un spectre plus diffus que celui effectué à proximité du périphérique : la dynamique sonore est plus faible, le volume également, il s'agit d'une zone presque calme. De plus, la plage de fréquences qui ressortent ici est nettement différente : il n'y a rien

de notable au-delà de 2000Hz sur le sonagramme A.6 alors que le sonagramme A.7 montre des données sonores lisibles au-dessus de 10000Hz.

En regardant plus en détails le drone du sonagramme A.6, on note deux nettes accentuations : une première aux alentours de 250Hz et une seconde autour de 1000Hz. Cette dernière semble correspondre au pic caractéristique du son de roulement (voir 2.2.1.2) causé par les pneus des véhicules sur l'asphalte. La représentation A.6 montre un drone perdant en intensité au-dessus de 1000Hz, deux bandes de fréquences sont cependant notables : une première dans l'extrême grave, légèrement sous la barre des 100Hz et une seconde autour de 250Hz.

Même atténué par la distance, le trafic environnant, ici essentiellement en provenance du Quai de Bercy, semble être la principale source du drone entendu sur le parvis de la BNF. Les accentuations dans les fréquences les plus graves semblent être celles qui persistent le mieux avec la distance, ces observations semblant recouper les analyses plus poussées décrites en 2.2.1.2.

3.3.1.4 États-Unis, San Francisco

Enfin, nous avons analysé trois enregistrements réalisés à San Francisco. Ces enregistrements se sont déroulés sur une journée avec un enregistreur portable Olympus LS-5.

Le sonagramme A.8 est un enregistrement effectué en centre ville, au niveau de la circulation ; le sonagramme A.9 est la représentation d'un enregistrement depuis le sommet d'un parking dans le centre ville (huitième étage) ; le sonagramme A.10 est la représentation d'un enregistrement depuis les hauteurs du parc Buena Vista, légèrement excentré du centre ville.

L'observation des trois sonagrammes fait ressortir que la dynamique et le volume de l'ensemble du paysage enregistré semblent diminuer avec l'éloignement de la circulation, ainsi que la largeur de la plage de fréquences captées. Des sons caractéristiques à cette ville ressortent nettement, on note la présence des sirènes sous forme d'oscillations entre 700 et 1400Hz ainsi que leurs harmoniques (A.10). Le passage d'un tramway, sonagramme A.8 autour de la minute une, offre un timbre et un motif rythmique inouïs dans les enregistrements européens précédemment présentés. Le sonagramme A.8 fait également ressortir ce qui peut s'apparenter au son de roulement du trafic, autour de 1000Hz tandis que, sur le sonagramme A.9, une nette accentuation

harmonique sur l'ensemble du fichier a lieu autour de 1660Hz. Cependant, ce n'est qu'en prenant une nette distance avec le centre ville que le drone urbain semble se détacher le plus finement : le sonagramme A.10 montre bien le tapis sonore généré par la ville sous 2000Hz. En regardant à nouveau les deux autres enregistrements, on note la présence d'accentuations persistantes, presque harmoniques, entre 20 et 2000Hz. Les événements sonores plus transitoires deviennent progressivement dominants au-dessus de cette limite.

3.3.1.5 Conclusion

L'ensemble de ces représentations spectrographiques urbaines nous montre des constantes, principalement dans le bas-medium et le grave du spectre. Il en ressort que chaque lieu enregistré semble posséder une signature sonore qui lui est propre et qui semble perdurer sur territoire acoustique étendu.

Cette étude sommaire mériterait un approfondissement avec des mesures plus rigoureuses dans le positionnement des points d'enregistrement (distances relatives, élévation) et dans le traitement des données recueillies : des analyses statistiques permettraient sans doute d'apporter des résultats plus fins et détaillés. Il serait également intéressant de cerner l'ensemble des variations de la partie discriminée comme étant l'arrière-plan en comparant des enregistrements répétés au même endroit.

Cependant, cette approche, même succincte, fait ressortir des drones s'apparentant à la résultante des mécanismes de diffusion acoustique en milieu urbain. Le recours à des points d'enregistrement distincts, pour un même paysage sonore urbain, montre des résultats similaires à ceux observés par Morgan [Mor09] lors de la collecte des matières sonores pour la construction des ambiances du jeu Prototype (voir 2.3.2) : l'enregistrement d'ambiances urbaines en des points éloignés de sources directes permet d'isoler le fond diffus de la ville, ces enregistrements peuvent ensuite être le point de départ de l'élaboration d'arrière-plans sonores dynamiques. La composante perceptive, que nous abordons dans les paragraphes suivants, paraît donc alors un élément important dans l'élaboration de paysages sonores urbains virtuels.

3.3.2 Le point de vue perceptif

Compte tenu de ce qui a été abordé dans les paragraphes précédents et de ce qui est connu du phénomène de non-localisation des réverbérations tardives, l'une des caractéristiques premières de l'arrière-plan sonore est son absence de localisation spatiale. La nature diffuse de l'arrière-plan lui donne une lisibilité unique dans un ensemble complexe : il se distingue des sources ponctuelles par son ubiquité. En règle générale, il n'est donc pas possible de distinguer précisément les uns des autres les objets qui composent l'arrière-plan : l'arrière-plan est un objet unitaire.

Cependant, si nous excluons l'éloignement géographique entre les sources et le point d'écoute, nous pouvons focaliser notre attention sur un objet sonore et l'isoler des autres sources qui s'en trouvent relayées ensemble à l'arrière-plan. Cet objet ainsi isolé par l'attention qui lui est portée change de plan (passe de l'arrière-plan au premier plan) sans changer de nature acoustique. Ceci est communément référencé sous l'effet dit *cocktail party* [Aro92] et nous amène à une autre caractéristique perceptive : notre attention ne peut être focalisée sur un objet d'arrière-plan. Un objet d'arrière-plan est quelque chose « que l'on entend sans l'écouter. » Par voie de conséquence la focalisation de l'attention sur un objet sonore renvoie en arrière-plan d'autres objets qui, d'un point de vue physique, ne sont pas nécessairement plus lointains. La manière dont l'écouter oriente son attention, sa perception, réorganise les frontières de l'arrière-plan.

Un autre phénomène à considérer est la nature des informations apportées par l'arrière-plan sonore urbain. Le premier aspect est d'ordre acoustique et concerne le type d'urbanisme environnant le point d'écoute. A l'oreille, une rue bordée de hauts immeubles en verre se distingue facilement d'une avenue Haussmannienne ou d'une zone pavillonnaire. La coloration spatiale du son est liée à la dimension et aux matériaux des surfaces avec lesquelles il entre en contact, le son renvoyé par la surface (sous forme de réflexion, de diffusion, de diffraction) ne conserve qu'une partie de son spectre : certaines fréquences sont absorbées et d'autres renvoyées. Ces données sont à mettre en relation avec l'effet canyon décrit plus haut (voir 2.2.1.1), leur ensemble fournissant une description relativement riche de la mécanique acoustique urbaine.

En plus de la coloration liée à la position de l'écouter, l'arrière-plan est partiellement significatif du contexte temporel et des condi-

tions météorologiques (vent, pluie, présence de la neige) [RV13]. Selon l'heure de la journée ou le jour de la semaine, certains axes urbains sont plus empruntés que d'autres, créant des mouvements plus ou moins réguliers dans la dynamique sonore de l'environnement urbain et colorant l'arrière-plan en fonction. Ponctuellement, les intempéries agissent comme des strates sonores qui viennent se mélanger au bruit de fond avec des effets de masquage ou de filtre. Précisons également que les conditions météorologiques influencent les utilisateurs de la ville dans leurs méthodes de déplacements (par exemple, une forte pluie incitera davantage à se déplacer en voiture plutôt qu'à pied), ce qui peut avoir des répercussions sur l'arrière-plan sonore.

Il est donc possible pour l'écouter de prendre conscience, perceptivement, de l'ambiance liée à une zone géographique : zone avec un fort trafic automobile ou piétonnier, présence d'un fleuve, d'animaux, d'un marché,... Le point de vue perceptif est fortement lié à la sensation de cohérence entre l'environnement et la composante sonore. Cela peut être relié à la notion de *présence* qui se comprend comme le sentiment d'être dans un environnement particulier alors que l'on se trouve, physiquement, dans un autre [SW98]. D'autre part, cette sensibilité perceptive de l'écouter est prise en compte par les designers sonores lors de l'élaboration de scènes virtuelles : dans la majorité des applications de type jeu vidéo, une scène sonore doit être perceptiblement crédible plutôt qu'être exacte du point de vue de l'acoustique.

Dans une acceptation commune et dans la pratique de la création audiovisuelle, l'ensemble de ces informations d'ambiance sont accentuées par des éléments ponctuels de signalétique urbaine. Ainsi, une ambiance nocturne de la ville de New York, d'un point de vue perceptif, peut se construire avec une sirène typique dans une acoustique fortement réverbérée. La signalétique urbaine et l'ambiance générée par les voix dans une foule ne correspondent pas totalement aux caractéristiques précédemment définies de l'arrière plan puisqu'il s'agit d'éléments identifiables, mais ils sont indispensables aux qualités perceptives d'un arrière-plan construit. Dans les paragraphes suivants, nous analysons donc le rôle de l'arrière-plan par rapport aux autres objets sonores d'une scène.

3.3.3 Le point de vue de la composition et de la mise en scène

Compte tenu de ces informations perceptives, lors de l'élaboration d'une scène, l'arrière-plan est l'objet abstrait qui va créer l'ambiance servant de support aux autres objets sonores de cette scène. Les sons d'avant-plan fournissent des informations sur l'action en cours et directement observable. Les sons ainsi mis en avant servent à guider l'attention de l'écoutant ou à illustrer son point de vue afin de renforcer la cohérence audiovisuelle.

La construction de l'objet arrière-plan peut se faire dans un premier temps à partir d'extrapolations des objets directement observables par l'utilisateur. Cependant, le rôle de l'arrière-plan sonore n'est pas d'illustrer l'arrière-plan visuel mais de fournir des informations sur l'état de l'environnement selon des axes préalablement définis : contexte (naturel ou urbain), densité moyenne de peuplement et nature du peuplement (humaine ou mécanique, par exemple), conditions climatiques, que les objets sources soient ou ne soient pas dans le champ de la caméra. En ce sens, un arrière-plan, dans une œuvre audiovisuelle, n'est pas nécessairement réaliste : certains éléments peuvent être exagérés pour augmenter ou atténuer le réalisme d'une scène. Les arrière-plans urbain restent cependant des éléments diégétiques : si les sources du contenu sonore ne peuvent pas être vues, elles peuvent être néanmoins déduites comme un prolongement logique de ce qui est directement observable.

Peuvent venir ensuite s'ajouter des éléments non-diégétiques tels que de la musique. Dans un jeu vidéo, la partie musicale se mélange à l'arrière-plan et appuie l'ambiance recherchée (anxiété, contemplation, temps accéléré, relâchement, etc.). L'arrière-plan se trouve porteur d'informations avec une esthétique qui peut lui être propre. Ses limites ne sont pas posées par un horizon précis, il peut représenter une perspective infinie. Le poids de l'imaginaire devient important dans cette représentation presque symbolique de l'état de la scène, le « regard » de l'arrière-plan se portant bien au-delà des bâtiments visibles.

Une telle construction répond à des critères perceptifs et non réalistes, l'arrière-plan n'est plus un prolongement acoustique de la scène mais un objet protéiforme, parfois lui même composite, à part entière.

L'arrière-plan sert également à maintenir une certaine « continuité spatiale » entre différentes sources ou différents espaces de la com-

position de sorte qu'il soit possible de lier les différents événements observables à un contexte, une ambiance, un environnement. Il peut s'agir également de fournir le liant acoustique aux différents éléments de la scène afin d'en renforcer le sentiment de cohérence et de crédibilité. En conséquence, il est possible de créer une discontinuité en surimposant un arrière-plan incohérent avec les éléments diégétiques². La dissonance ainsi créée est elle-même porteuse de sens dans la logique de la composition puisque reliant des éléments spatialement et/ou temporellement éloignés.

Esthétiquement, s'il fallait comparer la structuration de l'arrière-plan à celle d'une pièce de musique, celui-ci met en retrait des notions de rythme et de mélodie au profit de l'harmonie. Il s'agit d'une texture continue, diffuse, constante dans ses variations, évoquant un mélange de bruits filtrés et d'harmoniques dont il résulte une coloration spécifique à un lieu. Il n'est pas rare de pouvoir identifier une note fondamentale dans le spectre de l'arrière-plan et, à ce titre, Raymond Murray Schafer parle de l'arrière-plan comme d'un « bourdon », terme que l'on retrouve dans le langage de la composition musicale. L'arrière-plan est la fondation sur laquelle repose les autres objets sonores.

Ainsi que décrit précédemment, l'arrière-plan, dans sa masse, possède très peu de dynamique et celle-ci évolue lentement. Cependant, dans une logique de composition le séparant des autres objets, l'arrière-plan possède sa propre intensité. Le mixage d'un film rend souvent possible une balance plus ou moins forte des pistes d'ambiance ce qui permet d'accentuer ou mettre en retrait des éléments hors-champs. Dans les jeux vidéos, il est fréquent que le joueur puisse agir directement, par le biais d'un menu, sur la balance des différents éléments composant la bande son (musiques, ambiances, effets spéciaux, voix, sons d'interface, etc.). Il est ainsi donné la possibilité au joueur de mettre en avant ou en retrait tout ou partie de l'arrière-plan sonore.

La stylisation de l'arrière-plan peut aller jusqu'à l'abstraction. Ainsi, dans le film *Blade Runner* (1982), les plans les plus larges, montrant une anticipation de la ville de Los Angeles, sont illustrés par de la musique qu'il est sans doute plus simple de rendre évocatrice qu'un drone urbain. Un schéma identique de construction se retrouve régulièrement dans la série de films *Ghost in the Shell* (1995 et 2004), les nappes de synthétiseur remplaçant le drone acoustique dans les paysages d'anti-

2. Comme par exemple dans le film *Hiroshima mon amour* [Mor98]

cipation. Ce genre de stylisation brouille la frontière entre diégétique et non-diégétique des arrière-plans sonores.

3.3.4 Dynamique de l'arrière-plan

Considérer les plans sonores d'une scène d'un point de vue dynamique permet d'aborder les relations entre ces différents plans. Comme nous l'avons vu dans le Chapitre 2, les scènes interactives sont souvent segmentées en trois plans : un premier plan reproduisant des sources individuelles, un plan moyen composé de sources groupées et un arrière-plan constitué de pistes d'ambiance et/ou musicales. Les relations entre le premier et le plan moyen apparaissent clairement : un objet peut passer de l'un à l'autre, devenir individuel dans la scène, se mêler à un groupe, un groupe pouvant avoir des tailles et des propriétés différentes, des sources individuelles pouvant se regrouper.

3.3.4.1 Positionnement de l'écouter et changement de plan pour un objet sonore

Les relations avec l'arrière-plan, cependant, apparaissent moins évidentes. Contrairement aux autres plans, l'arrière-plan n'est pas spatialisé ni composé d'une somme d'objets clairement identifiables. L'arrière-plan n'est pas la somme sonore des objets observables multipliée par l'acoustique moyenne. Dans le monde réel, l'arrière-plan peut être composé en partie de sources sonores directement observables mais également d'objets bien plus lointains (comme un périphérique situé à plusieurs centaines de mètres du point d'observation), selon la position des frontières entre les plans.

Nous ne pouvons donc pas parler d'une continuité dynamique entre l'arrière-plan et les autres plans. L'arrière-plan n'est pas une simple diffusion des autres plans même s'il peut être partiellement caractérisé, du point de vue du timbre. Un objet ne passe pas toujours du premier plan à l'arrière-plan : les objets peuvent être caractérisés individuellement, peuvent caractériser des groupes mais peuvent également simplement disparaître s'ils sont trop loin pour être entendus. L'objet arrière-plan conserve son indépendance et ses propriétés propres.

Cependant, l'arrière-plan n'est pas un objet statique et ses transformations sont également liées à la position dynamique de l'écouter. En cela, l'arrière-plan devient un objet géographique et le point d'écoute

capte les propriétés du territoire qui l'environne. Tout déplacement du point d'écoute entraîne un réajustement du territoire perçu. Afin de rendre compte de cela, les techniques fréquemment utilisées dans le jeu vidéo ont recours au découpage du territoire en zones d'ambiance. L'écouter passe d'une construction sonore à une autre avec, en guise de transition, un fondu enchaîné (ou *crossfade*) plus ou moins important. Cette même démarche est utilisée pour les ambiances acoustiques par le biais de réverbérations de zone, à cette différence près que, si les frontières d'un espace acoustique sont définies par ses limites physiques, celles de l'arrière-plan sonores sont essentiellement perceptives.

Par la simple observation nous remarquons que dans le contexte d'une rue bondée, distinguer le pas d'un piéton parmi les autres peut nécessiter de se trouver à un ou deux mètres de celui-ci. Le même piéton considéré dans la même rue, une fois celle-ci déserte, sera entendu distinctement d'une distance bien plus lointaine. Dans les deux cas, un arrière-plan sonore existe, avec des frontières posées plus ou moins loin du point d'observation. Dans le premier cas, à quelques mètres de distance la masse sonore de la foule devient indistincte, compacte et compose un arrière-plan. Dans le second, ce seront probablement la circulation distante, la somme de sons de ventilations réverbérés et d'autres sources éparses qui formeront un arrière-plan plus « lointain ».

3.3.4.2 Dynamique de la composition

Les exemples du jeu vidéo nous montrent que le contenu des arrière-plans peut être plus varié et complexe que le bourdon décrit dans la théorie du Paysage Sonore. Définir les limites de l'arrière-plan revient donc en partie à délimiter son contenu.

Jusqu'à présent nous avons vu que l'arrière-plan peut inclure des sons diégétiques et non diégétiques, également des pistes musicales. Le cinéma apporte des éléments plus libres et détachés des préoccupations acoustiques. Si du point de vue acoustique il est possible de définir l'arrière-plan comme une somme de réflexions, le point de vue cinématographique offre une toute autre perspective : l'arrière-plan sonore d'une scène urbaine peut n'être composé, par exemple, que de bruits de pas, simplement de voix ou de sons naturels, occultant ainsi d'autres sons et le caractère réaliste de la scène. L'absence de réalisme n'exclut cependant pas la cohérence de l'ensemble de la scène, la composition

de l'arrière-plan suit des choix esthétiques et dramaturgiques, il s'agit de participer de l'atmosphère générale de la scène. Toutefois, l'arrière-plan constitue toujours un ensemble relativement monolithique, il est un support pour les sources émergentes de la scène appuyant ainsi une perspective plus ou moins profonde dans un champ plus ou moins large.

Ces positions cinématographiques plus abstraites se retrouvent moins dans les jeux vidéo qui privilégient, dans le cas de villes réalistes, une cohérence acoustique plus proche du monde réel (son spatialisé en trois dimensions, effets d'atténuation avec la distance, filtrages, réverbérations). Ceci s'explique par la continuité des plans dans l'expérience du joueur : d'une part, la position de la caméra cinématographique, de par son alternance, peut mettre en exergue des contrastes plus marqués que celle qui accompagne le joueur, d'autre part, dans le jeu vidéo, la caméra reste centrée sur le joueur et l'ensemble sonore doit donc suivre une cohérence dans le temps. Le jeu vidéo n'a pour autant pas recours à un rendu exhaustif de la scène, le point de vue perceptif est le plus prisé pour les scènes complexes et ce pour deux raisons principales : le besoin d'économies de ressources mémoire et processeur dans un premier temps et la volonté de donner à entendre une scène dont l'intelligibilité du contenu est parfaitement maîtrisée dans un second temps.

3.3.4.3 Arrière-plan et objets singuliers

Une sirène entendue au loin est beaucoup plus présente que le bourdon diffus du trafic et, surtout, attire bien davantage l'oreille. D'autre part, elle constitue une source identifiable contrairement au bourdon du trafic qui rend une impression des sources mélangées. Pourtant, ce son est tout aussi lointain et possède des caractéristiques en commun avec le bourdon en raison de la distance entre l'origine et l'écouter : la sensation de localisation est faible, la texture est réverbérée et filtrée, elle ne renvoie à aucune action directement observable. En elle-même, la sirène peut faire partie de l'arrière-plan puisqu'elle rend compte de caractéristiques descriptives de l'environnement. A cette différence près que, contrairement à ce que nous évoquions plus haut, la sirène pourrait passer d'un plan lointain à un plan rapproché et qu'elle évoque un objet singulier de l'environnement.

La question est donc de savoir si la sirène constitue un événement de la scène, c'est-à-dire si elle sert l'écouter à s'orienter, à se guider

dans l'action ou s'il s'agit d'une simple coloration servant à illustrer le tissu urbain, à créer une tension sans autre objet que l'atmosphère de la ville elle-même ou à renseigner sur la géographie de l'environnement.

Il peut en être de même pour tous types de sons clairement identifiables de loin, des explosions par exemple. Dans ce dernier cas, si la ville virtuelle illustrée est une zone de guerre, les explosions peuvent faire partie du tapis d'ambiance naturelle de la scène, décorréliées de toute action précise.

Un certain nombre d'objets sonores peuvent ainsi se mélanger aux textures d'arrière-plan. Ils prennent le statut de sons hors-champs *passifs*, c'est-à-dire que leur contenu ne doit pas guider l'attention de l'interacteur, contrairement aux sons hors-champs *actifs*, souvent utilisés dans les jeux vidéos pour guider le joueur dans la scène. La distinction perceptive entre les sons hors-champs *passifs* et *actifs*, soit la capacité d'attraction d'un son, dépend de nombreux facteurs tels l'intensité, le contenu spectral, l'évolution temporelle et la cohérence avec le contexte [Fin09], relativement aux autres sons de la scène.

3.4 Conclusion

La singularité de l'arrière-plan, considéré comme un objet sonore, et sa capacité à caractériser l'ambiance d'une scène en font un objet d'étude à plusieurs dimensions. Dans un premier temps, c'est un objet qui possède une grande variabilité de timbres relative à la fois à l'état de la ville à un moment donné mais également en partie à la position de l'écouter, ne serait-ce que du point de vue dynamique. Dans un second temps, le drone urbain possède une forte valeur perceptive dans la mesure où il contextualise naturellement les autres sons audibles en leur fournissant une « tonalité » de base.

3.4.1 Caractéristiques de l'arrière-plan sonore urbain

Considéré du point de vue spatial, l'arrière-plan urbain possède des frontières dynamiques relatives à la densité de la scène urbaine. Malgré cet aspect dynamique, il existe une sensation de continuité et d'enveloppement de la scène sonore. Cette stabilité semble inhérente à la nature de l'arrière-plan sonore.

À l'écoute d'enregistrements d'ambiances urbaines, il apparaît qu'un certain nombre de sons « passifs » et distincts (sirènes au loin, klaxons, sons de nature,...) viennent s'ajouter pour enrichir la texture continue du drone urbain et participent de la caractérisation du lieu. Une simple texture continue, malgré sa singularité spectrale, ne serait peut-être pas suffisante pour caractériser un lieu, son ambiance et son état.

Si l'on considère la création d'un synthétiseur sonore dédié, l'enjeu d'un tel outil est de faire ressortir l'ensemble des nuances des composantes de l'arrière-plan et leur capacité d'évocation, tout en conservant la cohérence avec le support visuel.

Ce chapitre a permis de mettre en exergue des caractéristiques essentielles des arrière-plans sonores urbains. Leur timbre complexe semble unique à chacune des villes que nous avons enregistrées. La mise en place d'un fond sonore crédible et interactif serait donc une étape non négligeable de la sonorisation d'une ville virtuelle ainsi qu'un atout immersif.

Prises séparément, la plupart des méthodes de (re-)synthèse que nous avons présentées dans le Chapitre 2 ont peu de chances de rendre compte de la complexité du spectre de l'environnement urbain, ou bien

la quantité de calculs nécessaire serait telle que l'implémentation dans un média interactif temps-réel en deviendrait plus difficile. Au regard des méthodes de synthèse sonore précédemment présentées et des caractéristiques des arrière-plans urbains, la synthèse granulaire, ou une méthode s'en approchant, semble la plus adaptée. Un certain nombre de paramètres devront cependant être adaptés aux besoins spécifiques des arrière-plans urbains, tant au niveau des paramètres de synthèse que de l'interface ou bien encore des effets à intégrer dans le parcours du signal.

3.4.2 Enregistrement de l'arrière-plan sonore urbain

Ce chapitre nous a permis également de définir, en lien avec le Chapitre 2, des modalités d'enregistrement des arrière-plans sonores urbain. Enregistrer des éléments lointain est une tâche particulière puisqu'elle semble requérir de se trouver à un endroit où seule est entendue la part diffuse et résiduelle du son de la ville (voir 3.3.1.5). Les principales difficultés sont donc d'éviter les sources à proximité des microphones (qu'il s'agisse de piétons, de moteurs, de vélos ou de sons naturels) et de saisir les nuances de texture dans l'arrière-plan.

Les points situés en hauteur ainsi qu'en périphérie des zones d'agglomération sont moins susceptibles d'être exposés directement aux sources sonores. Ce sont donc des endroits recherchés pour capter la coloration acoustique et le drone résiduel de la ville. Les zones calmes situées à l'intérieur de la ville, comme des parcs, sont également privilégiées car elles mettent la ville à distance tout en étant entourées de l'ambiance urbaine.

Toutefois, il est possible d'effectuer des captations plus directes. Les zones aux abords des périphériques peuvent présenter, selon les horaires, une forte densité de trafic. Ce phénomène est combiné à la nature très résonnante de l'habillage bétonné des axes de circulation. Il en résulte une sonorité très épaisse, réverbérée et très peu dynamique. Des enregistrements de ce type constituent des matières de départ efficaces pour recréer des atmosphères lointaines, notamment en les traitant avec des filtres.

Certains lieux proposent des ambiances qui varient au fil de la journée, de la semaine ou bien encore des conditions climatiques. Il s'avère donc utile d'y revenir à plusieurs reprises afin de capturer différentes

nuances d'une même ambiance. La pluralité des enregistrements s'avère également efficace en multipliant les points d'écoute d'une même scène. Enregistrer un lieu depuis plusieurs points d'écoute est un moyen d'obtenir un dégradé de nuances de ce lieu, que ce soit dans la densité perçue (par des enregistrements plus ou moins éloignés de certaines sources) ou dans sa coloration acoustique. Il s'agit donc de concevoir un synthétiseur permettant de naviguer dynamiquement entre les différentes atmosphères et de composer des nuances d'arrière-plan qui mélangent les différentes prises.

Ces principes que nous venons d'exposer sont à la base des enregistrements réalisés spécifiquement pour le synthétiseur d'ambiances urbaines que nous décrivons dans le chapitre ci-après. Les conditions techniques des enregistrements sont décrites dans la section 4.6.

Dans le chapitre suivant, nous présentons les principales étapes de la réalisation d'un synthétiseur dédié à la restitution d'arrière-plans sonores urbains ainsi que ses caractéristiques techniques.

Parmi les différentes méthodes de synthèse qui s'offrent à nous, la synthèse granulaire est une base crédible pour la réalisation d'un tel synthétiseur. Son utilisation permet de recourir à des enregistrements réels tout en laissant libre cours à l'implémentation d'autres outils de traitements du signal afin de procéder à de plus amples transformations du timbre.

Chapitre 4

Réalisation du synthétiseur

Sommaire

4.1	Introduction	125
4.2	Synthétiseur d’ambiances urbaines : objec- tifs et contraintes	126
4.2.1	Outil de design sonore	126
4.2.2	Fonctionnement en temps réel	127
4.2.3	Compatibilité avec Terra Dynamica	128
4.3	Pure Data	130
4.3.1	Historique et versions	130
4.3.2	Pure Data et interactivité	131
4.3.3	Pure Data comme environnement de déve- loppement	132
4.4	Architecture	133
4.4.1	Oscillateurs granulaires	137
4.4.2	Mixeur	142
4.5	Interfaces	145
4.5.1	Paramètres haut-niveau	145
4.5.2	Paramètres bas-niveau	146
4.5.3	Utilisation d’un contrôleur externe	148
4.6	Construction de la banque sonore de référé- rence	150
4.6.1	Matériel d’enregistrement	150
4.6.2	Matériel d’édition	151
4.7	Conclusion	153

4.1 Introduction

Ce chapitre est consacré aux différentes étapes de réalisation du synthétiseur d'arrière-plans urbains. Les différents aspects de la réalisation sont liés aux données, analyses et réflexions collectées au fil des chapitres précédents. Notre objectif est également de concevoir un outil utilisable par un designer sonore, à ce titre, il nous paraît essentiel que les fonctionnalités du synthétiseur ne soient pas en contradiction avec les méthodologies employées par les designers sonores.

Le synthétiseur était, dans un premier temps, destiné à être intégré au projet Terra Dynamica. En tant qu'environnement interactif temps-réel, le concept du projet Terra Dynamica a fixé une partie des objectifs et contraintes du synthétiseur que nous détaillons ensuite. La réalisation d'un tel outil de design sonore pose aussi des contraintes d'interface et d'édition de manière à rendre le design intuitif en mettant en avant des paramètres haut niveau perceptifs.

Après avoir exposé les principes de la méthode de synthèse que nous proposons de réaliser, nous présentons l'environnement de développement pour la réalisation du prototype (Pure Data) et l'architecture du synthétiseur. Cette architecture se décompose en plusieurs sous-ensembles organisés hiérarchiquement en plusieurs *patches* Pure Data.

Nous présentons ensuite l'interface du synthétiseur en détaillant ses paramètres haut-niveau et bas niveau. Cette section décrit également l'utilisation du synthétiseur avec un contrôleur physique permettant une édition rapide des principaux paramètres. Nous abordons enfin la manière dont ces paramètres sont liés au média temps réel pilotant la synthèse.

Le synthétiseur est alimenté par une banque de fichiers sonores dont les principes de réalisation sont exposés en 3.4.2. Nous présentons ici les modalités techniques de sa construction : choix du matériel d'enregistrement et réglages, édition des enregistrements.

4.2 Synthétiseur d'ambiances urbaines : objectifs et contraintes

Il n'existe à l'heure actuelle pas d'outil de design sonore spécifiquement dédié aux arrière-plans sonores pour les villes virtuelles. Pourtant, comme nous l'avons vu au cours du Chapitre 2, les arrière-plans sont pleinement inclus dans le paysage sonore de nombreux environnements virtuels interactifs. De nombreux aspects liés à la narration et aux ambiances sont retranscrits dans les arrière-plans. Nous proposons la conception et la réalisation d'un outil dédié à la génération dynamique des arrière-plans urbains.

Les choix dans le développement du prototype sont soumis à certaines contraintes. Il est important que le prototype soit suffisamment souple dans son architecture pour être capable de tester différentes méthodes de synthèse facilement sans remettre en cause l'ensemble de la structure. Cela implique de pouvoir implémenter rapidement une méthode de synthèse en particulier mais également de pouvoir faire fonctionner plusieurs de ces méthodes en parallèles, autant pour en comparer les résultats que pour générer un flux audio résultant de leur combinaison.

Cette section aborde les objectifs et contraintes liées aux fonctionnalités de design sonore, au fonctionnement du synthétiseur dans un environnement virtuel temps-réel et à la compatibilité avec le projet Terra Dynamica.

4.2.1 Outil de design sonore

Nous proposons donc de créer un outil qui donne au designer sonore les moyens de mettre en place des stratégies de sonification. Cet outil doit pouvoir répondre aux besoins du designer sonore en termes de place à donner aux arrière-plans dans la construction de la scène, de contenu sonore à intégrer aux arrière-plans et de méthodologie d'interactivité. Le synthétiseur que nous proposons de réaliser doit permettre au designer sonore de déterminer de quelle manière l'exploration de l'environnement et l'état de ce dernier modifient les arrière-plans sonores. L'ensemble de la synthèse doit fonctionner en temps réel et répondre aux manipulations d'un utilisateur dont on ne peut, de fait, prévoir le comportement.

Dans la conception de ce synthétiseur, notre intention est de donner au designer sonore des moyens d'élaborer des stratégies de sonification. Une attention particulière doit donc être portée à la flexibilité de l'architecture de sorte que le designer sonore puisse se l'approprier pour résoudre des problématiques liées à la gestion des arrière-plans sonores.

Notre objectif étant de créer un outil de design sonore, il a semblé important de réduire l'édition de la synthèse à quelques paramètres haut niveau. Cette intention vise à effectuer un maximum de modifications sur le son avec un minimum de manipulations et à offrir une interface simple, dans la lignée des synthétiseurs classiques. Indirectement, une interface graphique simple rend possible l'utilisation d'un contrôleur physique externe, plus intuitif à manier que la souris de l'ordinateur, et ce afin de tester des comportements sonores.

Le choix des paramètres est donc important ainsi que leur réactivité. L'architecture du synthétiseur est conçue de telle façon que certains paramètres haut-niveau peuvent modifier plusieurs paramètres bas-niveau simultanément. La description de l'action des paramètres haut-niveau n'est alors pas liée à son action sur la synthèse mais davantage à l'effet perceptif recherché par l'action de ce paramètre, tout en sachant que, en fonction du son sur lequel le synthétiseur agit, les affections perceptives peuvent varier.

4.2.2 Fonctionnement en temps réel

Une des principales contraintes pour la réalisation du synthétiseur est son fonctionnement en temps-réel. Cette contrainte se traduit sous deux aspects : la consommation d'espace mémoire et de temps de calcul.

Comme montré dans la section 2.4, les outils procéduraux de sonorisation peuvent consommer beaucoup de ressources en temps de calcul et les ambiances sonores sont le plus souvent composées de longues boucles audio utilisant beaucoup de mémoire.

Il s'agit donc ici de trouver un compromis entre les ressources mémoire nécessaires pour produire une variété de textures suffisantes et la quantité de traitements à appliquer sur les sons chargés en mémoire.

Un autre aspect important du fonctionnement en temps réel est la réactivité de la synthèse. Il convient de lisser les changements de valeurs pris en entrée du synthétiseur afin d'éviter de brusques variations qui donneraient un côté artificiel au rendu. Et, en même temps, il s'agit

d'assurer que les changements de valeur agissent sur la synthèse sans produire d'artefacts ou des résultats non souhaités.

4.2.3 Compatibilité avec Terra Dynamica

L'intégration du synthétiseur d'ambiances urbaines avec un environnement virtuel dynamique temps-réel est un des principaux enjeux de ce travail de thèse. Il s'agit de rendre effective la communication entre l'environnement et le synthétiseur afin de faire évoluer les paramètres de synthèse, et donc le rendu, en temps réel. D'autre part, le synthétiseur ne réalise pas l'intégralité de la sonorisation de Terra Dynamica. En effet, une partie du design sonore est réalisée dans le moteur audio Fmod. La contrainte résultante est d'insérer de façon cohérente le flux audio du synthétiseur à ceux produits par Fmod. Cette dualité crée le contexte pour valider l'efficacité du synthétiseur d'arrière-plans comme support sonore pour des sources individuelles.

Le synthétiseur est interfacé directement avec le noyau du logiciel Terra Dynamica. Le noyau envoie les informations relatives au peuplement de la scène virtuelle, notamment le type des acteurs, leurs positions et leurs chemins. Le noyau ne fournit cependant pas une description de l'environnement, c'est-à-dire qu'il ne met pas à disposition des paramètres généraux relatifs à l'ensemble de la scène. Le synthétiseur d'ambiances urbaines utilise donc des paramètres haut-niveau généraux qui requièrent l'acquisition de données qui englobent l'ensemble des acteurs et pas seulement leurs comportements individuels. Il convient ensuite d'organiser les données récupérées depuis le noyau afin de les mettre sous une forme significative pour les paramètres de synthèse. Les conditions de l'environnement produit par Terra Dynamica sont très variables, tant sur le nombre d'acteurs que sur leurs comportements et leurs catégories, il est alors nécessaire de mettre ces entrées à une échelle cohérente pour la description des arrière-plans sonores correspondants.

Enfin, il convient de prendre en compte les entrées fournies par l'utilisateur de Terra Dynamica. Dans Terra Dynamica, le principal mode d'interaction est le déplacement d'une caméra libre dans la scène en 3D. Le point d'écoute de la scène est positionné sur la caméra afin de renforcer la sensation d'immersion en accompagnant dynamiquement la navigation. Un certain nombre des paramètres de synthèse doivent

donc réagir au positionnement de la caméra, notamment l'élévation par rapport au sol.

4.3 Pure Data

Pure Data est un environnement de programmation graphique permettant de générer des contenus sonores et visuels ainsi que de procéder à des traitements complexes du signal en temps-réel. Pure Data traite essentiellement des données audio, MIDI ou au format OSC mais permet également de produire des rendus visuels sur la base de synthèse ou d'images¹.

Pure Data (abrégé « Pd ») est un langage de programmation par flux de données² permettant de créer graphiquement des logiciels appelés *patches*. Les différentes fonctions algorithmiques utilisées pour la création des *patches* sont représentées par des objets disposés sur l'écran par l'utilisateur. Les objets sont reliés entre eux par des câbles transportant les données d'un objet à un autre. Chaque objet effectue une tâche spécifique, il peut s'agir d'opérations mathématiques simples (additions, divisions,...) ou de fonctions audio et vidéo complexes telles que des transformées de Fourier, de la réverbération ou du décodage vidéo.

4.3.1 Historique et versions

Les bases de Pure Data sont posées en 1988 à l'Ircam (*Institut de recherche et coordination acoustique-musique*) par Miller Puckette, le logiciel s'appelant alors *The Patcher* [Puc88]. Cette première version fut revendue à la société Opcode (qui créera plus tard Max/Msp) tandis que Miller Puckette poursuivait le développement de la version *open source* du *Patcher* qui deviendra Pure Data.

Il existe aujourd'hui essentiellement deux versions de Pure Data : Pure Data *Vanilla*, version la plus sobre (et compilée sous forme de bibliothèque pour utilisation dans des environnements tiers), et Pure Data *Extended* qui inclut un choix plus large d'objets afin de réaliser plus facilement des opérations complexes qu'avec *Vanilla*.

Pure Data est couramment utilisé dans le cadre de créations artistiques, pour des applications scientifiques ou pédagogiques. Sa mise

1. De nombreux logiciels de programmation graphique permettant des rendus sonores complexes sont disponibles parmi lesquels nous pouvons citer Max/Msp (<http://cycling74.com/products/max/>), Reaktor, (<http://www.native-instruments.com/fr/products/komplete/synths-samplers/reaktor-5/>), Super Collider (<http://supercollider.sourceforge.net/>) ou Bidule (<http://www.plogue.com/products/bidule/>)

2. *Data flow programming language*

en œuvre simple qui consiste, *via* une interface graphique, à créer des *patches* en reliant entre eux des objets, chaque objet remplissant une fonction déterminée. Cette méthode de programmation, aujourd'hui partagée par d'autres logiciels et environnements de création, découle des expérimentations de Max Mathews dans les années 1950 qui ont par la suite inspiré le développement des premiers synthétiseurs analogiques modulaires.

L'ouvrage de référence proposé par Miller Puckette [Puc07] est une source importante d'informations quant aux possibilités de réalisation avec l'environnement Pure Data.

4.3.2 Pure Data et interactivité

Plus spécifiquement, Pure Data est utilisé dans de nombreux projets interactifs, le logiciel traitant en temps réel des variables reçues en entrée et restituant un flux sonore ou visuel en sortie. Les variables reçues en entrées peuvent provenir de manipulations directes de l'interface graphique de Pure Data, d'autres logiciels ou de surfaces de contrôles *via* la norme OSC (*Open System Control*), de protocoles réseau ou bien encore d'événements captés depuis l'environnement (mouvements, pressions, température, signaux, etc.) récupérés par le biais d'un dispositif matériel interfacé avec Pure Data³.

Pure Data peut également être intégré à des programmes plus complexes tels qu'un jeu vidéo. Il peut servir à des fins de prototypage pour tester rapidement des comportements sonores liés au moteur du jeu [Pau03] ou bien, comme dans le cas du jeu *Spore* être la base du moteur sonore. La versatilité des méthodes de synthèse et les possibilités d'optimisation [Far07c] de Pure Data en font un outil de développement efficace pour la sonorisation d'environnements virtuels.

Pure Data n'est pas limité à un registre de production ou de synthèse sonore. Son langage de programmation permet d'aborder de nombreuses méthodes de synthèse et de parvenir à des réalisations complexes. Il est possible de mettre en place des méthodes de lecture de fichiers sonores et de mixage plus avancées que celles disponibles dans les moteurs sonores de jeux vidéo. De plus, les *patches* sont imbricables les uns dans les autres, sous forme d'instance, ce qui permet de créer des ensembles riches. Les aspects logiques de la programmation dans Pure

3. Plusieurs dispositifs sont prévus à cet effet, notamment *Arduino*

Data sont comparables à ceux présents dans des langages tels que le C++, il est donc possible de contrôler finement le comportement d'un *patch* dans un environnement interactif.

Nous proposons donc d'utiliser Pure Data comme environnement de développement pour la réalisation d'un prototype de synthétiseur d'arrière-plans sonores urbains. Cette solution nous offre en effet des facilités de programmation et de développement tout en étant intégrable à des environnements virtuels.

4.3.3 Pure Data comme environnement de développement

L'ensemble des contraintes présentées précédemment soulignent le besoin d'une interface de programmation facile d'accès et flexible. La programmation graphique semble donc une méthode plus appropriée que la programmation classique, cette première nous offrant une meilleure visibilité de la dynamique des objets sonores programmés, de leurs interactions et des flux de données et audio qui les parcourent. L'incorporation d'éléments nouveaux, tels que l'ajout de modules de synthèse ou de traitements, et leur édition en sont simplifiées.

Les possibilités de la programmation graphique pour le design des environnements virtuels sont déjà exploitées pour la production des éléments visuels. Elles se développent également pour la programmation des attributs comportementaux. Du point de vue sonore, cette approche est assez peu explorée mais offre des perspectives prometteuses [Ven09].

La facilité d'exploitation et la versatilité ont motivé en partie notre choix d'utiliser Pure Data. La possibilité de convertir les *patches* en code par le biais d'une librairie (*LibPd*) offre la perspective d'une intégration complète du synthétiseur, au-delà de la programmation du prototype au format Pure Data. Les protocoles OSC et MIDI permettent de piloter facilement les *patches* à l'aide de contrôleurs physiques. Enfin, il est possible de faire circuler des informations entre Pure Data et un logiciel externe avec le protocole réseau UDP. La flexibilité et la modularité de Pure Data avec d'autres environnements de développements en font un outil efficace pour le prototypage et le développement de systèmes sonores interactifs [Pau03].

4.4 Architecture

L'architecture générale du synthétiseur est pensée comme celle d'un générateur de textures continues, tout en veillant à maintenir des variations et des évolutions dans le résultat sonore. Les variations et nuances sont à même d'améliorer la sensation de présence (d'immersion) de la trame sonore générée pour un environnement virtuel [FJ10].

Nous avons choisi d'utiliser la synthèse granulaire pour concevoir le prototype. Ce choix est motivé par plusieurs points.

Nous avons argumenté que les arrière-plans sonores urbains peuvent être considérés comme des textures sonores (voir 2.4.1.3), la littérature est explicite quant à la capacité des re-synthèses granulaires à produire des textures (c'est-à-dire la possibilité de générer des re-synthèses plus longues que les échantillons de départ) à partir d'enregistrements, ce qui nous conforte dans notre choix.

La littérature sur les villes virtuelles, et notamment le cas du jeu vidéo *Prototype* (voir 2.3.2) nous a montré qu'il est possible de reproduire des arrière-plans sonores urbains crédibles à partir d'enregistrement choisis. L'exploitation réfléchie des enregistrements semble donc, dans un premier temps, suffisante⁴. Nous restons dans cette ligne avec comme objectif de traiter les enregistrements comme des textures à re-synthétiser. Nous posons l'hypothèse que le choix d'une méthode granulaire de re-synthèse réduira l'espace mémoire requis par les échantillon avec en contre partie un faible coût supplémentaire en CPU.

Les méthodes granulaires impliquent donc de conserver l'étape d'enregistrement des sources. Nous argumentons, toujours dans le sens de la littérature, que cette étape fait partie intégrante du processus de design sonore. Plus que la simple collecte de données, les enregistrements sont aussi des sources d'étude et d'inspiration pour le designer [Mor09].

D'autre part, la réactivité dans le contrôle apporté sur les grains (tailles, déclenchements, enveloppes d'amplitude, etc.), particulièrement avec les méthodes de re-synthèse asynchrone (voir 2.4.1.3), peut être un facteur déterminant sur le rendu sonore. Techniquement, la

4. L'extraction par analyses des spécificités sonores caractérisant une ville afin d'en reproduire dynamiquement des ambiances à l'aide d'une méthode bas-niveau de re-synthèse pourrait cependant faire l'objet d'un autre sujet de recherche passionnant.

re-synthèse granulaire peut être suffisamment économe en ressources processeur pour être utilisée en temps réel dans des applications interactives ([Far07c], [Far07b], [Far10]). Nous proposons donc d'adapter la méthodologie de base de la re-synthèse granulaire aux besoins spécifiques de la génération de textures d'arrière-plans sonores urbains.

L'objet central du prototype (Fig.4.1) est un générateur de texture granulaire que nous avons nommé *Grainer*. Ces *Grainers* sont regroupés par quatre pour constituer ce que nous appelons des oscillateurs granulaires. L'ensemble des textures générées par les *Grainers* d'un oscillateur est exploré à l'aide d'un paramètre, à la manière d'une table d'ondes. Nous avons choisi d'intégrer trois oscillateurs au prototype, chaque oscillateur possède donc une table d'ondes à quatre entrées, suivie de filtres passe-haut et passe-bas.

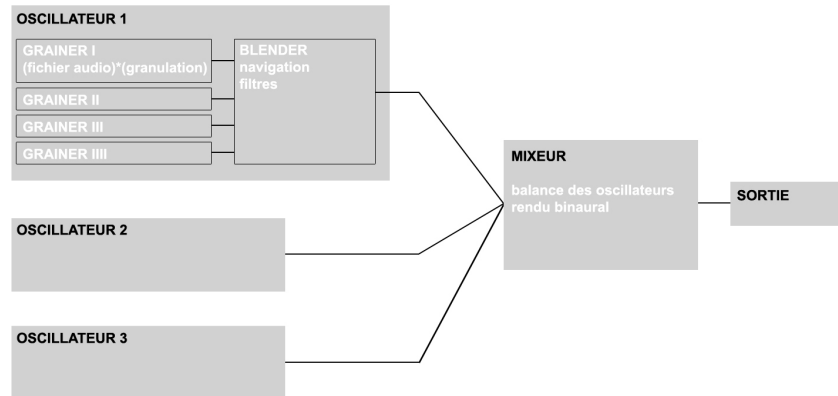


FIGURE 4.1 – Architecture simplifiée du prototype de synthétiseur : de gauche à droite, les oscillateurs composés de quatre *Grainers* dont le mélange et le filtrage sont contrôlés par le *Blender*, les oscillateurs sont ensuite mélangés dans le Mixeur principal.

Ces principes généraux donnent à l'utilisateur la possibilité de créer et manipuler des timbres environnementaux en superposant différents mélanges en temps-réel. Cette approche est courante dans la pratique du design sonore. Un enregistrement, quel qu'il soit, est souvent une approximation du fait sonore réel. De plus, aussi précis que puisse être l'enregistrement, il arrive souvent que les caractéristiques d'un son unique ne parviennent pas à produire une sensation suffisamment

riche et puissante pour appuyer l'image. Ainsi, les designers sonores ont souvent recours à une méthode de création par strates en superposant différents enregistrements ou sons de synthèse afin de produire un objet sonore. Par exemple, pour sonoriser un verre qui se brise, il est courant d'utiliser plusieurs sons de verre brisé, chacun sur une plage de fréquences différente (medium, aigüe,...) et d'appuyer l'ensemble avec un son synthétique ou acoustique plus grave. Le son ainsi artificiellement créé possède une plage de fréquences plus large et/ou plus contrôlée qu'un enregistrement seul. Il sera également plus simple de mélanger l'objet sonore ainsi créé avec le reste de la bande son : le spectre riche et contrôlé laisse à l'ingénieur de mixage la liberté de « creuser » certaines fréquences pour faire de la place aux autres sons simultanés ou bien mettre ce son en premier plan ou bien encore en retrait sans qu'il ne perde en intelligibilité.

L'architecture du synthétiseur que nous proposons reprend cette méthodologie en mettant à disposition plusieurs niveaux de mixage afin de contrôler le contenu et la densité de chaque strate de texture sonore. Cette architecture peut également être mise à profit pour mixer des textures hétérogènes, soit pour créer des arrière-plans complexes composés de plusieurs sous-ensemble, soit pour contrôler le mixage dynamique de différentes zones de l'environnement virtuel depuis une seule instance du synthétiseur. Il s'agit ici de proposition d'usages basées sur l'architecture du synthétiseur, au regard des différentes méthodes de sonorisation décrites dans le Chapitre 2. Cependant, nous souhaitons l'architecture suffisamment ouverte pour qu'il soit possible au designer sonore de se l'approprier et de parvenir à ses propres stratégies.

4.4.0.1 Exposé théorique de notre approche granulaire

Le but de notre utilisation de la synthèse granulaire est d'obtenir des matières sonores d'arrière-plan larges, denses, à partir de quelques enregistrements et contrôlables avec quelques paramètres.

En raison de la morphologie des arrière-plans sonores urbains, nous privilégions l'utilisation de grains de taille suffisamment grande (entre 100 millisecondes et plusieurs secondes) pour maintenir la sensation de continuité en suivant la pulsation lente, implicite, des enregistrements urbains lointains. Cette approche nous amène à envisager le fenêtrage dynamique des grains selon une méthode différente de celle habituelle-

ment utilisée (consistant à appliquer une enveloppe sur l'ensemble du grain, voir 2.4.1.3). Notre approche propose de distinguer deux enveloppes : une appliquée au début de chaque grain (un fondu entrant) et une appliquée à la fin de chaque grain (un fondu sortant). Alors que les fenêtres des méthodes décrites en 2.4.1.3 s'appliquent à l'ensemble du grain, notre approche nous permet de déterminer quelle proportion du grain est affectée par des variations de volume. Nous pensons que notre méthode permet de contrôler plus finement le comportement de la synthèse granulaire avec des grains de grande taille, particulièrement en ce qui concerne le séquençage des grains et leur enchainement. La suite de la démarche consiste à synchroniser les fondus afin de préserver la continuité de la texture. À cette fin, le choix de courbes de type *puissance égale* nous a paru donner les meilleurs résultats. Notre approche nous permet donc de contrôler la taille des grains ainsi que le format de l'enveloppe qui leur est appliquée.

Les processus du design sonore consistent souvent à superposer différentes strates sonores pour parvenir au résultat voulu (voir 4.1). Nous souhaitons intégrer cette méthodologie au synthétiseur afin d'offrir d'avantage de contrôle sur le rendu. Nous proposons de réaliser plusieurs granulations (*Grainers*) en parallèle et de mélanger dynamiquement les textures produites par chacune d'elles. Si plusieurs *Grainers* peuvent ainsi être superposés afin de créer une texture plus complexe, nous pensons que cela rend aussi possible le mixage entre des textures associées à différentes zones d'un environnement virtuel, par exemple. L'idée est ici d'offrir des possibilités à l'utilisateur (le designer sonore) sans le contraindre par une méthodologie prédéfinie.

Les différentes expérimentations que nous avons menées nous ont amené à envisager l'arrière-plan sonore urbain comme étant lui-même un ensemble complexe qu'il est possible de diviser en différentes strates de textures. En effet, le drone du trafic apparaît comme une texture elle-même complexe, variant selon les heures et les points d'écoute, mais à cette strate de base peuvent s'ajouter d'autres sons environnementaux comme la rumeur de la mer ou divers sons de nature et climatiques ou bien encore des voix. De plus, dans le contexte d'un média interactif, il paraît nécessaire de pouvoir passer d'une texture d'ambiance à une autre, de façon continue, sans rupture. La méthode la plus fréquemment utilisée consiste à faire se superposer les différentes ambiances, le temps de la transition. Pour ces différentes raisons, la

réalisation du synthétiseur s'est progressivement orientée vers la génération de textures multiples en simultané afin de contrôler des rendus complexes à l'aide d'un mixage dynamique.

Les *Grainers* sont initialement conçus selon des propriétés de synthèse granulaire synchrone. La superposition des différents flux de grains avec des réglages différents pour chaque *Grainer* entraîne des variations et des irrégularités dans le mixage perçu. Ceci tend à rapprocher le résultat de la synthèse granulaire asynchrone et à limiter la probabilité d'une sensation de répétition dans la texture produite. Notons également qu'il est possible, dans Pure Data, de rajouter des facteurs d'aléatoire sur l'évolution des valeurs afin de créer davantage de diversité.

L'ensemble de l'architecture du synthétiseur est conçue de manière à pouvoir créer un rendu narratif sur la base des enregistrements sélectionnés. Les méthodes d'enregistrements décrites en 3.4.2 sont mises en pratique en tenant compte du fonctionnement du synthétiseur. Le rendu final de la synthèse est la combinaison du choix des enregistrements et des choix d'utilisation du synthétiseur. Ainsi, des enregistrements multiples d'un même lieu, disposés dans les *Grainers* d'un oscillateurs, permettent de naviguer dynamiquement entre les différentes atmosphères et de composer des nuances d'arrière-plan qui mélangent les différentes prises. La structuration de l'approche narrative est laissée libre au designer sonore et différentes stratégies de sonification sont possibles. Notre proposition ne constitue pas une solution unique à la gestion des arrière-plans dans un environnement aussi complexe que peut l'être une ville virtuelle interactive. Le développement de chaque environnement virtuel pose ses propres contraintes, notre objectif est donc de concevoir un outil flexible et adaptable tant aux contraintes du développement qu'à la stratégie choisie par le designer sonore.

Dans les paragraphes suivants, nous détaillerons les deux principales sections de l'architecture exposée dans la Fig. 4.1 : les oscillateurs granulaires avec leurs composantes (*Grainer* et *Blender*) et le mixeur général du synthétiseur.

4.4.1 Oscillateurs granulaires

La synthèse granulaire permet d'obtenir des textures changeantes à partir des enregistrements réalisés au préalable. Le regroupement de

Grainers en oscillateurs granulaires est au centre de l'architecture du prototype.

Chaque oscillateur se divise en deux sous-parties dont nous décrivons les propriétés et fonction : quatre *Grainers* sont mélangés et filtrés dans un objet *Blender*.

4.4.1.1 *Grainer* - Lecteurs de grains

La partie granulaire du synthétiseur s'articule autour de séries d'objets *Grainers* (Fig. 4.2). A chaque *Grainer* est assigné un fichier audio. Ce fichier est sélectionné pour ses propriétés en termes de contenu, de dynamique et son registre fréquentiel, par exemple. Le choix et l'édition de l'enregistrement se font en amont de la lecture.

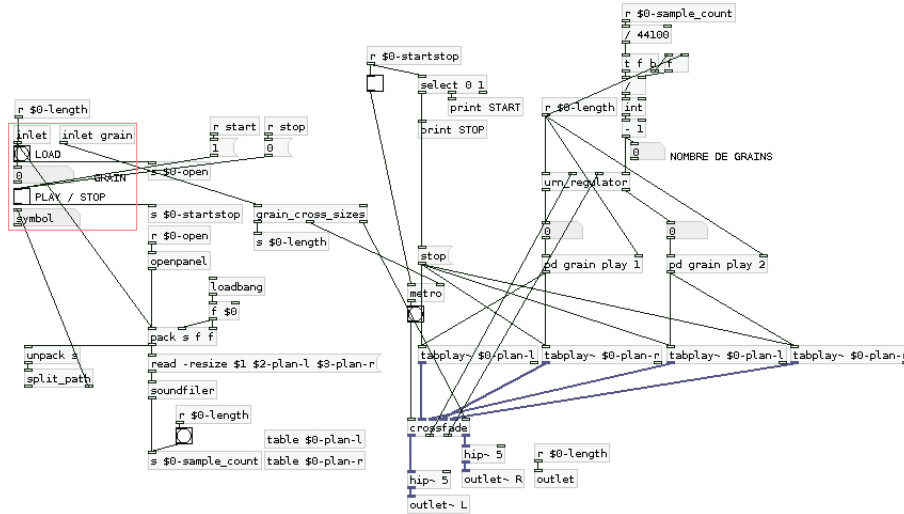


FIGURE 4.2 – Détails du *patch Grainer*.

Chaque *Grainer* travaille donc sur un fichier sonore qui lui est assigné. La taille de grain choisie détermine le nombre (arrondi) de grains exploitable par un *Grainer* : si le fichier dure 60 secondes et que la taille de grains est fixée à 5 secondes, le *Grainer* aura à disposition 30 grains (60 divisé par 5). La taille des grains est un paramètre dynamique, l'architecture prend donc en compte d'éventuelles modification de ce ratio en temps réel. Le nombre de grains disponibles est normalisé de manière à ce que chacun d'entre eux soit lié à une position précise dans le fichier. Un système d'urne (ou *shuffle*) détermine aléatoirement quel

grain est lu. Une fois la liste épuisée, l'urne retourne à son statut initial et recommence la sélection aléatoire. Ce système permet de minimiser la répétition des grains.

Si l'on observe les caractéristiques précédemment énoncées, le *Grainer* fonctionne en partie comme un synthétiseur granulaire à deux voix : deux grains issus d'un même fichier sont lus en parallèle. La cadence à laquelle les grains sont choisis dans l'urne est fonction de deux paramètres : la taille des grains et le temps de fondu enchaîné entre deux grains. Les fondus enchaînés entre les grains servent à éviter les artefacts de lecture numérique et, dans le cas présent, à masquer toute discontinuité de volume entre deux grains. Leur durée correspond à un ratio de la taille de grain sélectionnée (Fig. 4.3) ainsi, la modification de la taille de grain affecte également la taille des fondus entre les grains tout en préservant les proportions entre les deux valeurs.

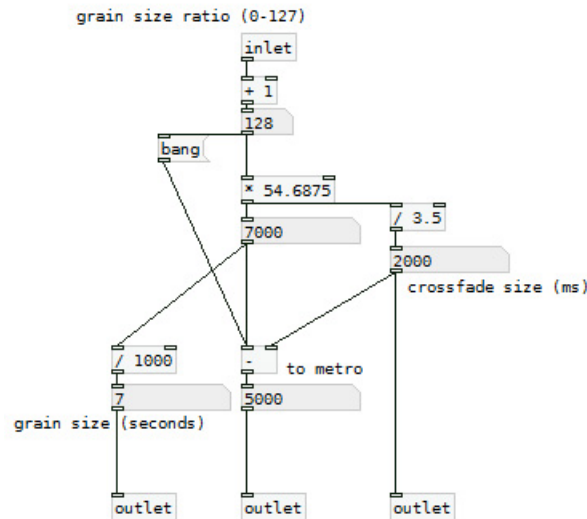


FIGURE 4.3 – Détails du *Grainer*, ici le sous-*patch* servant à calculer le ratio entre la taille de grain et la taille des fondus.

La taille des grains varie de 50 millisecondes à 7 secondes. Nous avons choisi de privilégier les grains de grande taille car ils préservent l'intégrité de la texture enregistrée. Les essais que nous avons réalisés avec des grains d'une taille inférieure à deux-cent millisecondes font ressortir, en juxtaposant des fragments distants du fichier original, les variations de matière. La rythmicité imprimée par la granulation est

trop disjointe de celle présente originellement dans le fichier enregistré. La taille des grains est relative à la source. En effet, ceux-ci doivent être suffisamment longs pour que la source soit reconnaissable mais aussi suffisamment courts pour éviter la sensation de répétition. Le choix de la taille des grains affecte la texture du rendu, pouvant créer des effets de périodicité ou altérer la nature de l'enregistrement de départ. Il convient donc souvent d'ajuster manuellement la taille des grains en fonction de la source afin de trouver le bon équilibre. La taille des grains peut cependant ne pas être fixe et varier dans une certaine fenêtre temporelle afin de prévenir la survenue d'effets de périodicité.

Notons également que l'assignation d'un fichier sonore à un *Grainer* est un processus dynamique. Le fichier sonore est chargé dans la mémoire vive de la machine. Le *patch* conçu de telle sorte qu'il est possible, à n'importe quel moment, d'assigner un nouveau fichier sonore à un *Grainer*, ce second fichier remplace alors le premier dans la mémoire.

Nous avons choisi d'utiliser quatre instances de l'objet *Grainer* pour constituer un oscillateur granulaire. Le sous-*patch* *Blender* contient les paramètres nécessaire à l'association de ces quatre *Grainers* en un ensemble dynamique et cohérent.

4.4.1.2 *Blender* - Navigation et filtrage

Quatre lecteurs de grains (*Grainers*) sont assignés à un oscillateur. Les quatre granulations se font en parallèle, c'est dans le sous-*patch* *Blender* (Fig. 4.4) que les flux audio se mélangent selon deux paramètres décrits ci-après. Le flux audio résultant du mélange passe ensuite au travers de filtres.

Nous présentons d'abord la section du *Blender* dédiée au mixage des *Grainers* puis les filtres.

Mélange - *Navigation* et *Blend Factor*

Le mélange audio entre les *Grainers* d'un oscillateur s'apparente à une forme de navigation linéaire inspirée des paramètres des tables d'ondes (voir 2.4.1.2). Cette dernière s'effectue en ajustant une position relative aux différents *Grainers*. Cela permet de parcourir les différents lecteurs, du premier au quatrième, afin de choisir la texture qui convient. La transition entre deux *Grainers* peut être ajustée afin

de créer un effet ressemblant à un fondu enchaîné. Plusieurs *Grainers* peuvent être entendus simultanément selon le réglage des deux paramètres détaillés ci-après.

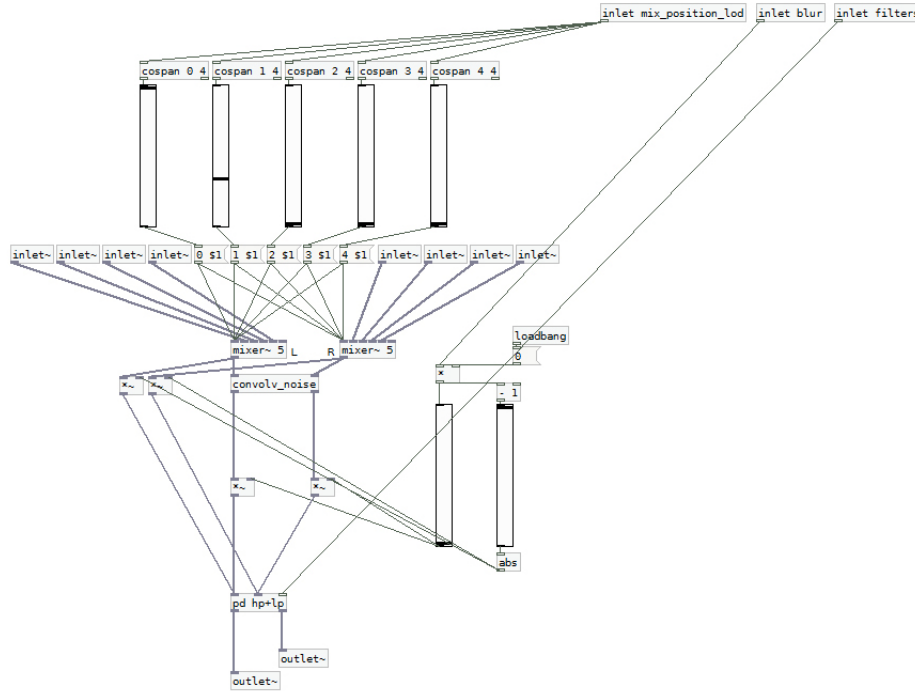


FIGURE 4.4 – *Blender* permettant de parcourir les différents *Grainers* d'un oscillateur

L'ajustement de cette position relative (paramètre *Navigation*) fait varier le volume des quatre *Grainers* simultanément. Le paramètre *Navigation* correspond au point central d'une gaussienne distribuant les valeurs de volume entre les quatre *Grainers*. L'ajustement de cette gaussienne est contrôlé par un second paramètre, le *Blend Factor*, qui détermine la largeur de la distribution des valeurs de la gaussienne. Le *Blend Factor* constitue une sorte de focale. Dans sa distribution la plus large, le *Blend Factor* mélangera le flux provenant des quatre oscillateurs et, dans sa distribution la plus étroite, il permet d'isoler un *Grainer* et de réduire les trois autres au silence.

Ces deux paramètres (*Navigation* et *Blend Factor*), utilisés conjointement, servent donc à modéliser la texture sonore en sélectionnant le *Grainer* principalement entendu tout en déterminant la proportion dans laquelle les autres *Grainers* de l'oscillateur sont audibles.

Filtres

Chaque oscillateur possède une paire de filtres, un passe-haut et un passe-bas. Ces filtres traitent en temps réel le mélange audio issu du *Blender*.

La combinaison des deux filtres permet de donner une impression de distance, d'éloignement ou d'occlusion par rapport au point d'écoute mais aussi de supprimer des fréquences inutiles à la composition de la texture d'arrière-plan. Ils peuvent être utilisés comme des outils de coloration du signal ou pour créer des effets rudimentaire d'acoustique. Le choix de ces filtres correspond à des critères perceptifs. En s'éloignant d'une source, le point d'écoute perd la perception des fréquences les plus aiguës et les plus graves.

L'implémentation de ces filtres est un moyen de rendre plus lointain un enregistrement qui comporterait des détails inappropriés à la sensation de distance recherchée et, de fait, offre la possibilité, en faisant varier dynamiquement la quantité de filtrage, de varier les sensations de proximité et de densité de la texture.

La combinaison du *Blender* et des filtres permet de créer des variations de dynamique et de texture complexes. Ces paramètres sont au centre des possibilités du synthétiseur. Il est ainsi, par exemple, possible de créer des effets de niveaux de détails en tenant compte de la distance du point d'écoute.

Les oscillateurs granulaires sont ensuite mélangés dans la section Mixeur du synthétiseur. Nous détaillons cette section dans les paragraphes suivants.

4.4.2 Mixeur

Le mélange et les réglages de niveau des oscillateurs se fait dans la section Mixeur du synthétiseur. Cette section intègre également deux modules de rendu au format Ambisonique.

4.4.2.1 Mixage des oscillateurs

La section mixeur consiste à mélanger les flux audio des instances des oscillateurs décrits dans les paragraphes précédents. Le volume de chacun des oscillateurs est contrôlable indépendamment par l'utilisa-

teur. Le volume des oscillateurs constitue également un point d'entrée intéressant pour des applications de mixage dynamique.

Après le *Blender*, le Mixeur est le deuxième étage de contrôle de volume à l'intérieur du synthétiseur. Si les contrôles du volume offerts par le *Blender* peuvent servir à contrôler l'évolution et le contenu d'une texture, le contrôle de volume des différents oscillateurs peut servir à régir le mixage des différentes textures entre elles.

L'application la plus simple consisterait à assigner chaque oscillateur à une zone d'une ville virtuelle, le mixage temps-réel entre les oscillateurs devenant alors relatif à la position du point d'écoute. Il serait également possible d'utiliser plusieurs oscillateurs pour créer une texture très complexe, le deuxième étage de mixage que constitue le Mixeur étant alors un paramètre supplémentaire dans le modelage du rendu.

4.4.2.2 Rendu au format Ambisonique

L'ensemble du prototype est conçu pour fonctionner au format stéréo. Cependant, le caractère immersif des textures d'arrière-plan peut nécessiter un rendu plus complexe (voir 2.4).

À cette fin, nous avons choisi d'intégrer une option permettant le rendu au format Ambisonique du flux audio produit par le synthétiseur.

Présentation de la technologie Ambisonique

La technologie Ambisonique [MM95] est théorisée pour la première fois par Gerzon en 1970 [Gue70]. Cette technique a déjà été comparée favorablement [JLP99] à d'autres méthodes de rendu sonore tridimensionnelle.

Une étude de 2007 [GLCB07] montre que, sur le plan perceptif, le format Ambinonique, comparé au stéréo et au transaural, apporte une meilleure sensation d'enveloppement et d'immersion. En contrepartie, l'écoute s'avère moins précise dans la localisation des sources et la lisibilité spatiale de la scène. Ces aspects négatifs ne présentent pas de contrainte particulière pour nous puisque l'essentiel des sons que nous produisons sont des textures très faiblement ou aucunement localisées dans l'espace de l'auditeur. Il est même bienvenu que l'espace de l'arrière-plan ne vienne pas concurrencer le positionnement plus précis des sources de premier plan.

Implémentation

Nous utilisons deux instances de l'objet `[+binaural~]`⁵ en sortie de l'étape de mixage des oscillateurs. Une instance est dédiée au canal droit et l'autre au canal gauche. Des informations d'angle sont apportées à chacune des instances avec un écartement de cent quatre-vingts degrés entre les deux canaux.

L'utilisation de ces objets permet d'effectuer une rotation de l'ensemble du rendu en liant le paramètre d'azimut du point d'écoute. La stéréo est recomposée en additionnant les sorties droite et gauche respectives des objets `[+binaural~]`. Le résultat donne la sensation de distribuer la scène sonore tout autour du point d'écoute. Cela est particulièrement perceptible en effectuant une rotation, certains éléments de la texture sonore semblent rattachés à un point virtuel de l'espace.

Si, du point de vue de l'immersion, cette technique peut donner des résultats probants, elle a tout de même un impact sur le rendu, en créant notamment des effets de phase. L'utilisation des objets `[+binaural~]` est donc un compromis dépendant surtout des besoins de la scène à sonoriser.

5. Cet objet fait partie de la librairie SoundHack créée par Tom Erbe et portée sous Pure Data avec les contributions de William Brent, Daniel Arias, Bryan Oczkowski et Trevor LeVieux, <http://puredata.info/downloads/soundhack>

4.5 Interfaces

Afin de rendre le synthétiseur facilement utilisable, les principaux paramètres de l'architecture sont mis à disposition du designer sonore au travers de deux niveaux. Les paramètres sont choisis et organisés afin de simplifier l'édition de la synthèse. Ces paramètres contrôlent chacun un ou plusieurs aspects des modules de synthèse.

Dans les paragraphes suivants, nous décrivons six paramètres permettant la modification du rendu de la synthèse. Ces six paramètres sont ajustables pour chaque oscillateur granulaire.

Les six paramètres se divisent en deux catégories. La première catégorie est constituée des paramètres haut-niveau. Il s'agit de descriptions de phénomènes psychoacoustiques plus complexes modifiés par la synthèse : leurs noms correspondent au ressenti recherché (par exemple « distance » ou « densité ») mais leur action modifie simultanément plusieurs aspects de la synthèse (filtres, échantillons lus, etc.) dont ils deviennent des abstractions. La seconde catégorie regroupe des paramètres bas-niveau. Ceux-ci affectent directement un paramètre de synthèse : ce qui est décrit par le paramètre correspond à ce qui est édité (par exemple la taille d'un échantillon lu).

4.5.1 Paramètres haut-niveau

Il s'agit des paramètres psychoacoustiques. Peu de paramètres se trouvent avoir une description haut niveau dans la mesure où leur impact sur le rendu est davantage lié à la combinaison de leur usage par l'utilisateur.

4.5.1.1 Atténuation

Ce paramètre fait varier la fréquence de coupure des filtres passe haut et passe bas, en parallèle. La diminution des fréquences aigües et des fréquences les plus graves renvoie à une sensation de distance accrue entre le point d'écoute et l'objet sonore. Les filtres sont réglés de sorte que l'atténuation est plus forte sur les hautes fréquences que sur les basses fréquences, en conformité avec les propriétés de diffusion acoustique urbaine. Cet étalonnage de ces courbes suit les descriptions des propriétés fréquentielles du bruit de fond urbain (voir 2.2.1.2).

4.5.1.2 Granularité

La modification de ce paramètre, pour chaque oscillateur, influence plusieurs aspects fondamentaux de la synthèse. Ainsi que décrit précédemment, la durée des fondus enchaînés est proportionnelle à la taille des grains (Fig. 4.3). Modifier la taille des grains revient à modifier également la taille des fondus enchaînés qui les mélangent entre eux. La taille est à ajuster en fonction des enregistrements. La taille optimale peut varier d'un enregistrement à un autre en fonction de son contenu.

Ce paramètre agit donc également sur l'objet `[metro]` qui se trouve au cœur de chaque lecteur. Cet objet détermine la cadence à laquelle les fragments de fichier audio sont sélectionnés par les objets `[tabplay~]` chargés de les lire.

La taille de grain optimale est à ajuster en fonction de chaque enregistrement. La granularité peut créer une rythmicité implicite au rendu, ce qui n'est pas forcément souhaitable dans ce cas d'une texture sonore. Si ce paramètre est donc principalement fonctionnel, des tailles de grains extrêmement réduites peuvent être utilisées à des fins de design sonore pour des rendus non-réalistes ou des effets de design sonore.

4.5.1.3 Position dans la table d'ondes

Ce paramètre est ajustable pour chaque oscillateur et se trouve avec les paramètres haut-niveau. Il est au cœur des possibilités de design sonore du synthétiseur. En lui-même, il ne correspond à aucune abstraction psychoacoustique si ce n'est celle que le designer sonore décide de concevoir dans l'organisation des fichiers chargés dans les *Grainers*.

Si ce paramètre correspond donc directement à l'action de mixage à l'intérieur d'un *Blender*, il peut donner l'impression de variations complexes, évoquant parfois un effet de *morphing* entre les différents sons.

4.5.2 Paramètres bas-niveau

Il s'agit des paramètres descriptifs affectant directement un paramètre unique de synthèse : ce qui est décrit par le paramètre correspond à ce qui est édité.

Ces paramètres ne sont pas forcément accessibles sur l'interface principale d'édition. Ils agissent, pour certains, davantage comme des réglages de calibration pour les paramètres haut-niveau.

4.5.2.1 *Blend Factor*

Il s'agit d'un paramètre complémentaire à la position dans la table d'ondes. Il permet d'ajuster la forme de la gaussienne contrôlant de le volume des différents *Grainers* d'un oscillateur.

Sa valeur maximale crée une gaussienne large mélangeant plusieurs *Grainers* simultanément ; sa valeur minimale isole un *Grainer* en particulier. La largeur de la gaussienne influe donc également sur la fluidité du mixage et la couleur des timbres produits.

4.5.2.2 *Ratio des fondus enchaînés entre les grains*

Ce paramètre est accessible à l'intérieur de chaque lecteur de grain (ou *Grainer*). Il est normalement influencé par le choix de la taille des grains. Cependant, le ratio peut être ajusté manuellement afin de contrôler plus avant la superposition des deux fenêtres de lecture.

Ainsi, la durée du *crossfade* peut, au maximum, presque égaler celle des grains ou, au minimum, devenir presque inexistante. Le calibrage de ce paramètre est à penser en fonction du type de textures synthétisées. Des textures très aléatoires et discontinues (comme la pluie ou d'autres phénomènes naturels) peuvent mieux fonctionner avec des *crossfades* courts tandis que des textures continues peuvent mieux fonctionner avec des temps de recouvrement plus longs.

4.5.2.3 *Calibration des filtres*

Les filtres passe-haut et passe-bas, synchronisés et réglables par l'intermédiaire du paramètre haut-niveau « Atténuation » peuvent être calibrés manuellement en bas-niveau afin d'ajuster leur effet sur le rendu.

Il s'agit d'ajuster la valeur maximum de filtrage (en Hertz) selon le paramètre d'atténuation, et ce pour chaque filtre.

Les trois paramètres haut-niveau sont modifiables en temps réel tandis que les trois paramètres bas-niveau constituent des points de

réglages qui ne sont initialement pas conçus pour être édités en temps-réel.

La modification en temps-réel des paramètres haut-niveau donne au synthétiseur son aspect dynamique. Ces modifications peuvent se faire manuellement sur l'interface graphique de Pure Data. Cependant, la manipulation à la souris de l'interface de Pure Data ne permet d'éditer qu'un paramètre à la fois. Or, les modifications de rendu les plus intéressantes peuvent requérir la manipulation de plusieurs paramètres simultanément. Cette étape de test permet d'effectuer des réglages avant d'intégrer le synthétiseur à un programme externe dont les variables peuvent modifier plusieurs paramètres de synthèse simultanément (voir 5.1).

Dans les paragraphes suivants nous décrivons l'utilisation d'un contrôleur Midi externe afin de pouvoir éditer les paramètres du synthétiseur en temps-réel.

4.5.3 Utilisation d'un contrôleur externe

La virtualisation des synthétiseurs et autres outils de design sonore n'en réduit pas moins l'importance de l'interface physique avec laquelle il est possible d'agir sur le logiciel. Les ingénieurs et designers sonores ont donc souvent recours à ce type d'interfaces, au format MIDI, USB ou OSC, pour créer, éditer et mixer des contenus sonores. L'interface physique permet de palier l'une des principales faiblesses de la manipulation à la souris : il devient ici possible d'agir sur plusieurs paramètres simultanément et d'avoir un retour plus tangible des éditions. Il a donc paru important de lier certains paramètres de la synthèse à une surface de contrôle physique. C'était également un moyen de retrouver une approche intuitive de la synthèse par la création et la transformation de timbres en temps réel par l'utilisateur.

Au fur et à mesure des tests conduits sur le synthétiseur, il est apparu que les transformations de timbre les plus efficaces requièrent le plus souvent la variation de plusieurs paramètres simultanément. Cela permet d'accentuer le naturel et la richesse des transformations en apportant plusieurs informations en même temps à l'écouter. Nous avons essentiellement utilisé un contrôleur qui envoie et reçoit des messages au format MIDI. Dans ce cas, la configuration de Pure Data requiert l'assignation d'un *Control Change* au paramètre qu'il va piloter. Cependant, l'utilisation du format OSC permet de contrôler Pure Data

avec des tablettes, *smartphones* et tout autre appareil compatible ce qui accroît fortement le nombre d'interfaces physiques envisageable pour interagir avec la synthèse.

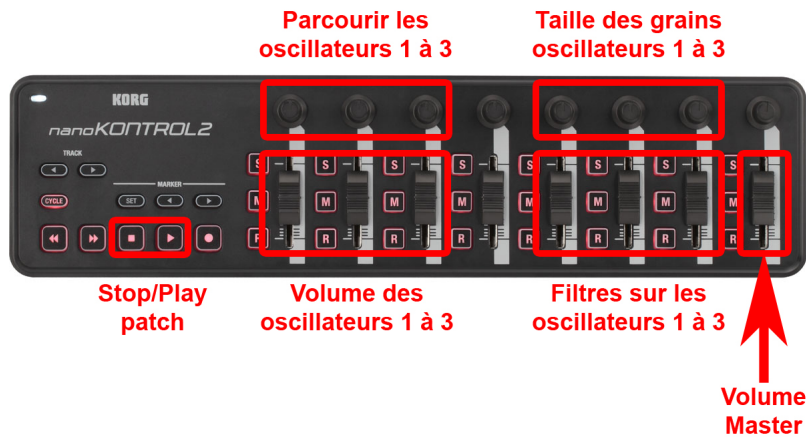


FIGURE 4.5 – Répartition des paramètres du synthétiseur sur un contrôleur Korg nanoKONTROL2

Le choix du contrôleur utilisé s'est porté sur le Korg nanoKontrol2, principalement en raison de sa portabilité et du nombre de contrôles correspondant globalement à ceux présents sur l'interface du synthétiseur. La Fig.4.5 présente l'interface du contrôleur ainsi que les paramètres de synthèse qui lui sont assignés.

L'utilisation d'un contrôleur a permis de tester le synthétiseur en faisant varier plusieurs paramètres simultanément. Pour l'utilisateur, cette interface est plus proche d'une intégration du synthétiseur à un environnement virtuel que ne l'est la manipulation à la souris. La manipulation de plusieurs paramètres simultanément permet de tester l'interaction des paramètres les uns avec les autres ainsi que de contrôler que des variations complexes ne créent pas d'artefacts dans le rendu de la synthèse.

4.6 Construction de la banque sonore de référence

Ainsi que décrit dans les Chapitres précédents, l'enregistrement de l'environnement urbain fait partie intégrante de la démarche d'analyse et de compréhension de la ville sonore.

Selon les principes exposés en 3.4.2, les enregistrements sont également un moyen de récupérer des données utilisables pour le synthétiseur. Dans les paragraphes suivants, nous détaillons le matériel que nous avons utilisé pour les réaliser ainsi que leur édition et intégration dans le synthétiseur.

4.6.1 Matériel d'enregistrement

Le matériel privilégié pour la collection de sources sonores dédiées au synthétiseur est un enregistreur portable (Olympus LS-5, voir Fig.4.6). Ce type d'enregistreur est fréquemment utilisé par les designers sonores du fait sa taille réduite et de sa facilité d'utilisation. Il est équipé d'une paire de microphones cardioïdes orientés à 90 degrés.



FIGURE 4.6 – Enregistreur portable Olympus LS-5

L'ensemble des enregistrements utilisés avec le synthétiseur sont réalisés sur la base de réglages identiques. Les enregistrements sont effectués au format .wav dans une résolution de 48 kHz et 24 bit. L'enregistreur possède les trois paramètres de réglages d'enregistrement suivants. Un interrupteur fait varier la sensibilité des microphones entre «*high*» et «*low*». Après différents essais, la position «*low*» semble donner les meilleurs résultats en termes de rapport signal/bruit. Un second interrupteur active ou désactive un filtre coupe-bas, nous le laissons désactivé afin de préserver au maximum la plage de fréquences captée.

Enfin, un encodeur détermine le gain de pré-amplification du signal de 0 à 10. Pour tous les enregistrements cet encodeur est sur 4, ce réglage semble présenter le meilleur rapport signal/bruit pour les enregistrements que nous effectuons. Enfin, les enregistrements se déroulant en extérieur, une bonnette est positionnée systématiquement sur la paire de microphones afin de les préserver au maximum du vent.

La méthode présentée ici est celle utilisée pour réaliser les enregistrements qui ont servi à tester et à intégrer le synthétiseur. Cependant, avec des moyens du temps supplémentaires, il aurait été intéressant d'effectuer des essais avec d'autres techniques et formats d'enregistrement. Le choix de l'enregistreur stéréo portable est avant tout pratique car, en se suffisant à lui-même, il ne requiert pas l'installation d'un dispositif complexe. La contrepartie de ce système portable est l'absence de contrôle sur les différents éléments de la chaîne d'enregistrement : choix des microphones, pré-amplificateurs et convertisseurs, positionnements de microphones. Les microphones de l'enregistreur utilisé sont de type cardioïdes, or d'autres enregistreurs (comme le Sony PCM-D50) proposent des microphones de type omnidirectionnel (c'est-à-dire qu'ils captent de façon égale les sons tout autour). Il aurait été intéressant aussi de procéder à des enregistrements au format binaural, Pure Data étant capable de décoder ce type de format. Un futur travail devrait être consacré à la comparaison de différents formats d'enregistrement et à leur éventuel impact sur le rendu perceptif du synthétiseur d'ambiance.

4.6.2 Matériel d'édition

Les enregistrements sont ensuite édités afin de s'assurer de n'en conserver que les fragments dépourvus de signaux parasites (sons paraissant trop proches ou donnant l'impression d'être hors contexte). Ce travail d'édition s'est fait dans le logiciel Logic Pro.

Outre le nettoyage des prises, il convient, si besoin, de mettre les enregistrements à des niveaux cohérents les uns avec les autres. Cette tâche, s'apparentant à du design sonore peut aller à l'encontre de la perception des événements sonores dans les conditions réelles : de manière à ce que les enregistrements puissent être perçus de façon audible et crédible dans un mixage interactif, il faut parfois en modifier le contenu et augmenter ou diminuer au préalable leur volume sonore.

Cet étalonnage a aussi pour but de faciliter, par anticipation, le mixage de la partie d'arrière-plan avec les sources spatialisées.

Il convient également parfois aussi d'atténuer la présence de certaines fréquences à l'aide d'un égaliseur paramétrique précis. Cette étape peut permettre de supprimer des constantes sonores (moteurs, voix, souffle, entre autres) indésirables sur l'enregistrement.

Ces étapes d'édition nous permettent de constituer des banques sonores cohérentes et fonctionnant au mieux ensemble. Ces altérations aux enregistrements, parfois subtiles, peuvent aider à fluidifier les transitions entre les grains et à fluidifier le mixage temps-réel des ambiances produites. L'automatisation de ce processus serait coûteuse en temps de calcul et en développement et priverait le résultat d'un facteur humain propre à chaque designer sonore.

4.7 Conclusion

Dans ce chapitre, nous avons défini les principes de création d'un outil de synthèse d'arrière-plans sonores urbains pour un environnement virtuel interactif temps-réel et audio visuel. L'architecture de synthétiseur que nous avons proposé est la résultante tant d'une approche théorique et empirique de la notion d'arrière-plan sonore urbain que des contraintes de réalisation d'un outil de design sonore temps-réel pour les villes virtuelles. L'analyse détaillée de l'architecture mettant en application ces principes est accompagnée d'un mode d'emploi opérationnel partant de l'enregistrement des échantillons allant jusqu'à la définition de stratégies de sonification correspondant à différents usages. L'Annexe B propose des exemples d'utilisation plus détaillés pour l'utilisation du synthétiseur.

Notre intention était d'utiliser des paradigmes issus de la synthèse granulaire pour répondre au besoin complexe de la création de textures dynamiques. Pour cela, nous avons privilégié la superposition et le mixage temps réel de lecteurs granulaires fonctionnant en parallèle. Cette méthode, combinée à des filtres soustractifs, constitue une base pour créer des arrière-plans complexes, variées et dynamiques.

Le synthétiseur créé dans Pure Data est constitué de trois parties principales :

- *Grainer* : lecteur granulaire générant une texture continue sur la base d'un fichier audio qui lui est assigné
- Oscillateur : un ensemble de quatre *Grainers* mixés et filtrés dynamiquement
- Mixeur : mélange et traite les flux audio de trois oscillateurs granulaires

Au cours de la réalisation du prototype, nous avons expérimenté l'utilisation d'autres méthodes de synthèse ainsi que l'ajout de *patch* complémentaire. Cependant, nous trouvions que cela complexifiait l'interface et focalisait l'utilisation sur des choix technologiques plutôt que sur des stratégies de sonorisation.

Nous avons présenté le choix des paramètres haut-niveau et bas-niveaux exposés pour manipuler le synthétiseur. L'organisation de ces paramètres est essentielle à la manipulation de la synthèse par un utilisateur. Nous avons proposé d'interfacer ces paramètres avec un contrôleur physique afin de parvenir à des manipulations plus complexes.

Notre objectif était de parvenir à produire un outil de synthèse granulaire suffisamment réactif pour rendre possible la création d'une palette sonore intéressante à partir de quelques fichiers sonores, manipulés par un nombre réduit de paramètres. Dans le chapitre suivant, nous présentons l'intégration du synthétiseur au projet de ville virtuelle Terra Dynamica ainsi qu'une première expérimentation pour la validation perceptive de la synthèse sonore proposée.

Chapitre 5

Validation et expérimentations

Sommaire

5.1	Intégration au projet Terra Dynamica . . .	157
5.1.1	Conception sonore pour le projet	157
5.1.2	Réalisation	160
5.1.3	Bilan et conclusion	169
5.2	Écoutes comparatives	170
5.2.1	Expérimentation	171
5.2.2	Présentation et analyse des résultats	174
5.2.3	Discussion et conclusion	177

Ce Chapitre présente deux expérimentations réalisées avec le synthétiseur décrit précédemment. La bonne intégration du synthétiseur dans un environnement virtuel ainsi que sa confrontation avec d'autres méthodes de (re-)synthèse nous semblent en effet indispensables à la validation de son fonctionnement. Les aspects techniques du synthétiseur sont testés (telles que la robustesse de la méthode de synthèse ou la réactivité aux variations d'un environnement), ainsi que des aspects plus esthétiques ou relevant de capacités de production en termes de design sonore.

L'ensemble de ce travail de thèse s'est déroulé en parallèle du développement du projet Terra Dynamica. Ce projet a fourni un cadre pour intégrer et tester les différentes fonctionnalités du synthétiseur. Si certains aspects du projet Terra Dynamica ont agi comme des contraintes sur le développement du synthétiseur, nous avons cherché à garder une architecture la plus ouverte possible afin de pouvoir aisément transposer les concepts fondamentaux du synthétiseur à la sonorisation d'autres villes virtuelles.

L'intégration du synthétiseur d'ambiances dans un environnement virtuel ne nous semblait pas suffire à valider l'entièreté de la démarche. Nous avons donc choisi de mener une expérimentation basée sur des écoutes comparatives pour tenter de montrer que le processus de synthèse que nous avons mis en place n'est pas moins efficace qu'un autre pour la gestion de textures sonores d'arrière-plans urbains.

La première section de ce chapitre concerne donc l'intégration du synthétiseur au projet de ville virtuelle Terra Dynamica. Dans ce projet, le synthétiseur est en charge de la sonorisation des arrière-plans sonores en temps-réel.

La seconde section décrit une expérimentation visant à valider, par des écoutes comparatives, l'efficacité de l'objet *Grainer* en tant que générateur de textures sonores.

5.1 Intégration au projet Terra Dynamica

L'objectif général du projet Terra Dynamica¹ est d'animer les données collectées lors des projets Terra antérieurs (topographie de Paris, reliefs des façades ou encore marquages au sol) afin de créer une ville virtuelle dynamique régie par noyau, soit un système d'intelligence artificielle.

Le domaine d'application de Terra Dynamica n'est pas défini de façon fermée. Le développement du projet se fait en perspective de démonstrateurs thématiques : sécurité et sûreté, transports ou encore applications citoyennes. Chacun de ces démonstrateurs utilise à sa façon le noyau de la simulation gérant le déplacement dynamique des acteurs leurs états ainsi que ceux de la simulation. L'un des principaux enjeux de la simulation est donc d'être réactive à une grande diversité de scénarios et d'usages.

La Fig.5.1 montre le schéma d'ensemble du projet par briques de contributions technologiques. La partie la plus bas-niveau, au centre, concerne la géométrie de l'environnement, viennent ensuite se greffer les briques dynamiques de la simulation gérant la navigation, l'intelligence artificielle ou encore les modèles physiques, puis les briques plus haut-niveau fournissant les rendus sonores et visuels.

Le CNAM intervient à différents niveaux de la réalisation de Terra Dynamica : réalisation d'une installation artistique (*Le Promeneur Écoutant* [LP13]), de la brique chorégraphie de foules [LNT12] et de la sonorisation 3D.

Dans cette section de la thèse, nous nous intéressons à la réalisation du moteur d'animation sonore 3D et à l'intégration du synthétiseur décrit au Chapitre 3. Nous verrons dans un premier temps les besoins exprimés pour la sonorisation de Terra Dynamica. Ensuite, nous détaillerons la solution mise en œuvre avec les implications respectives de Fmod et du synthétiseur ainsi que leur fonctionnement avec la simulation. Enfin, nous dresserons un bref bilan de l'intégration du synthétiseur dans le projet et de son interaction sonore avec Fmod.

5.1.1 Conception sonore pour le projet

Les objectifs de la sonorisation étaient de parvenir à un rendu sonore crédible et immersif mais qui soit également capable de dispenser

1. www.terradynamica.com

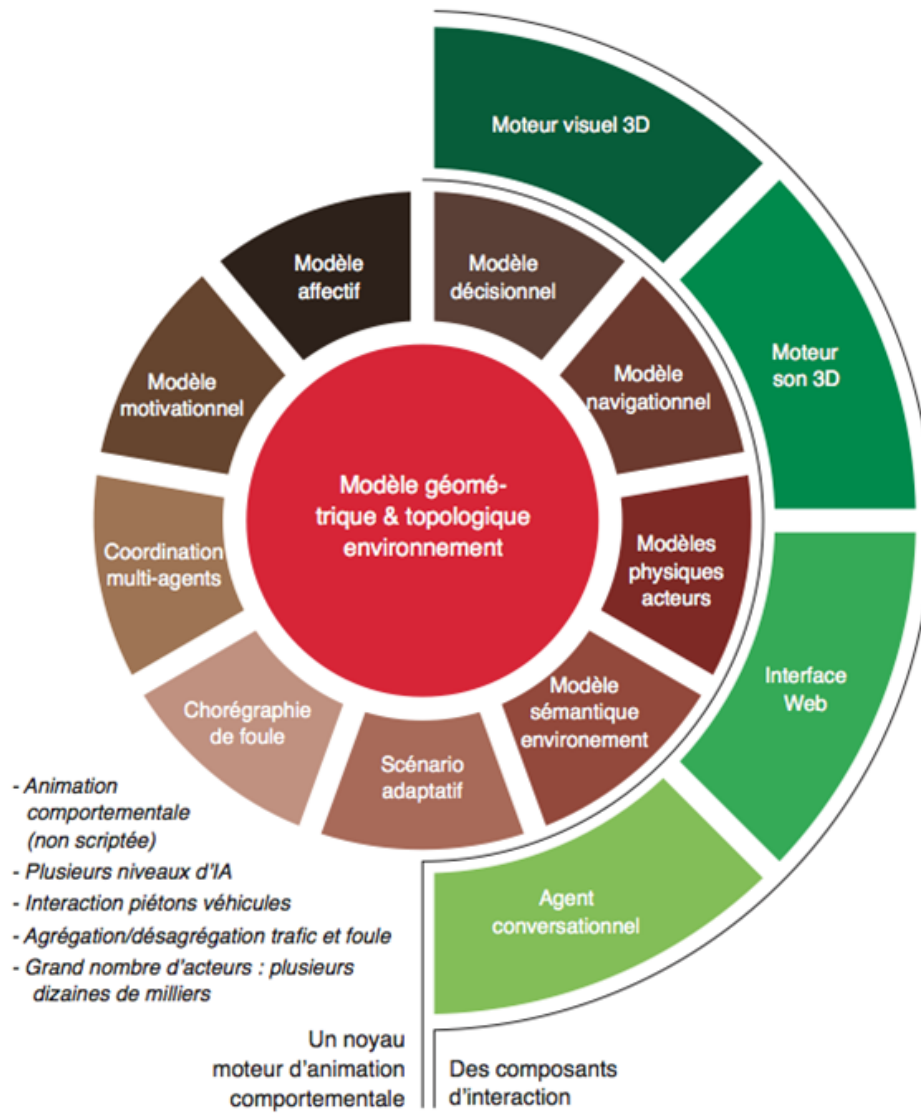


FIGURE 5.1 – Schéma d'ensemble de la conception du projet Terra Dynamica

à l'utilisateur des informations qui ne peuvent être obtenues par le retour visuel, telles que la sonorisation de sources hors-champs.

Les contraintes pour la sonorisation du projet Terra Dynamica concernent principalement la composition des scènes. En effet, dans le cadre de la simulation, une scène virtuelle peut contenir jusqu'à

plusieurs milliers d'acteurs. Ces sources sonores sont réparties sur de grandes distances (le démonstrateur sur lequel portait l'essentiel des tests réalisés consistait en un arrondissement complet de Paris), la sonorisation de chacune de ces sources en continu pose des contraintes techniques (nombre de fichiers audio lus simultanément, calcul de spatialisation pour chaque source) qui nécessitent de choisir quelles sources sonoriser à quel moment et selon quels paramètres. Comme nous l'avons vu dans le Chapitre 2, les scènes urbaines complexes sont souvent sonorisées en segmentant spatialement et sémantiquement les sources autour du point d'écoute : les sources individuelles les plus pertinentes sont maintenues tandis que les autres sont progressivement agrégées dans la distance jusqu'à un arrière-plan diffus et ubiquitaire. La nature dynamique de cette segmentation permet de faire ressortir ou de cacher une source selon les besoins de la scène en s'appuyant sur l'effet de masquage entre les sources. Notamment pour des raisons de performance, nous décrivons dans les paragraphes suivant une stratégie de sonorisation inspirée de celles utilisées dans le jeu vidéo pour la gestion des sources de Terra Dynamica.

Terra Dynamica présente deux classes de sources sonores : les piétons et les véhicules. Les piétons possèdent des sons de déplacement c'est-à-dire des bruits de pas correspondant à la marche. Les véhicules sont sonorisés pour leur déplacement (moteur et roulement) ainsi que pour deux événements (avertisseurs et freinages brusques). Les véhicules se déclinent en quatre catégories : voitures, camionnettes, camions et bus. Les types de piétons sont au nombre de trois : femme adulte homme adulte et enfant.

Afin de maintenir la diversité et la cohérence, chaque type d'acteur doit posséder au moins une série de sons de déplacement qui lui sont propres. Certains sons sont également spécifiques à un type de source, par exemple le klaxon et l'avertisseur (cloche) du bus sont caractéristiques de cette source uniquement afin de la rendre clairement identifiable dans le paysage sonore.

Une fois les sources sonores définies et catégorisées, une des principales difficultés de la sonorisation évoquée plus haut concerne la densité des scènes. Nous proposons donc une stratégie particulière à Terra Dynamica afin de rendre des scènes sonores avec un grand nombre d'acteurs dynamiques.

5.1.2 Réalisation

La base de la stratégie de sonorisation pour Terra Dynamica consiste à répartir les champs sonores entre le moteur audio Fmod² (voir 2.4.2.1) et le synthétiseur granulaire d’ambiances décrit dans le chapitre précédent.

Les outils de production de Fmod sont efficaces pour la sonorisation et la spatialisation dynamiques de sources. Le synthétiseur proposé génère des fonds d’ambiances ubiquitaires et dynamiques. La stratégie que nous proposons consiste donc à articuler une frontière acoustique virtuelle entre Fmod et le synthétiseur d’ambiances. Fmod est chargé de la sonorisation des sources individuelles reposant sur les strates d’ambiances du synthétiseur qui fournissent une perspective acoustique à l’ensemble de la scène en même temps qu’un liant entre les sons spatialisés.

La Figure 5.2 est une schématisation de la répartition du rendu sonore entre Fmod et le synthétiseur réalisé dans Pure Data. Fmod et Pure Data reçoivent tout deux des informations depuis le noyau de la simulation. Fmod sonorise individuellement les sources de premier plan tandis que Pure Data recrée des arrière-plans procéduraux sur la base des données transmises par la simulation.

Dans les paragraphes suivants, nous détaillons les utilisations respectives de Fmod et du synthétiseur d’ambiances urbaines dans le cadre de Terra Dynamica.

5.1.2.1 Utilisation de Fmod

Fmod est une solution audio couramment utilisée pour la sonorisation de jeux vidéo et d’environnements interactifs complexes. Les fonctionnalités proposées sont nombreuses, nous utilisons ici principalement celles de lecture et de spatialisation de fichiers sonores ainsi que le module *Engine Designer* que nous détaillons un peu plus loin.

La spatialisation des sources est une fonctionnalité essentielle dans le rendu sonore de Terra Dynamica puisqu’elle permet d’animer le déplacement sonore d’un piéton ou d’un véhicule dans l’espace de la scène

2. Fmod est un moteur audio dédié à la production d’environnements sonores interactifs tels que les jeux vidéo ou les installations. Il s’agit d’une librairie programmable et d’une série d’outils permettant de définir les caractéristiques de lecture de fichiers sonores. Pour la sonorisation de Terra Dynamica, nous avons utilisé la version 4.38.03. www.fmod.org

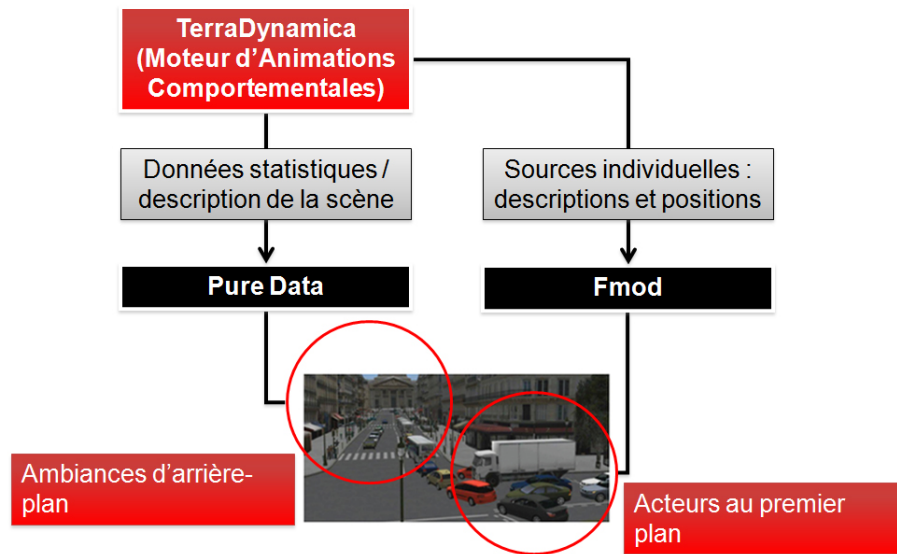


FIGURE 5.2 – Sonorisation de Terra Dynamica : schématisation de l'articulation dynamique entre Fmod (sources individuelles) et le synthétiseur réalisé dans Pure Data (arrière-plans), présentation du type de données transmises de la simulation vers Fmod et Pure Data

virtuelle. Les principaux paramètres de la spatialisation définissent la portée de la source (la distance sur laquelle la source est audible) et la forme de sa diffusion (une sphère pouvant être affinée en un cône directionnel). La Fig.5.3 montre les principaux paramètres de configuration d'une source audio 3D, incluant l'effet Doppler permettant de rendre plus réaliste le déplacement des sources autour du point d'écoute. Ces paramètres se situent dans l'interface haut-niveau de Fmod, servant principalement à régler le comportement de la lecture des sons en fonction de variables entrantes ainsi que les principaux effets appliqués. Ces paramètres sont appliqués aux sources individuelles (piétons et véhicules) de la scène.

Sonorisation des piétons

La sonorisation des piétons consiste à déclencher des bruits de pas lorsque Fmod reçoit l'information que l'animation de marche est activée sur un personnage. Les bruits de pas sont enregistrés, sélectionnés et édités dans un séquenceur audio³ puis intégrés dans Fmod. Lorsque

3. Dans le cadre de ce projet nous avons utilisé Logic 9.

Mode	3d
Ignore Geometry	No
3D Rolloff	Inverse
3D Min Distance	1
3D Max Distance	10000
3D Auto Distance Filtering	On
3D Auto Distance Center Freq	1500
3D Position	World relative
3D Min Position Randomization	0
3D Max Position Randomization	0
3D Cone Inside Angle	360
3D Cone Outside Angle	360
3D Cone Outside Volume	0
3D Doppler Factor	1
3D Speaker Spread	0
3D Pan Level	1
2D Speaker L	0
2D Speaker C	0
2D Speaker R	0
2D Speaker LR	0
2D Speaker RR	0
2D Speaker LS	0
2D Speaker RS	0
Speaker LFE	0

FIGURE 5.3 – Fmod : configuration des paramètres 3D d’une source sonore.

l’animation de marche est active, le sons de pas sont sélectionnés aléatoirement dans une liste attribuée au type de personnage sonorisé. Une variation aléatoire de hauteur (lecture plus ou moins rapide de l’échantillon) est appliquée à chaque lecture afin d’accentuer la diversité. Une dizaine de piétons est ainsi créé et répartie entre les trois types de personnages.

Sonorisation des véhicules

La sonorisation des véhicules comprend les sons de déplacement (moteur, roulement et aérodynamique) et d’événements (klaxon et/ou avertisseurs, freinage brusque). Les sons de moteur sont reproduits en faisant un mixage dynamique de courtes boucles audio du moteur enregistré à différents régimes tours/minute, tout en prenant en compte si le moteur est en charge ou non afin d’ajouter du réalisme. Pour les besoins du démonstrateur, deux types de surface de roulement sont distingués pour les véhicules : asphalte et pavés. Le son de roulement sur chacune de ces surfaces est reconstitué à partir d’enregistrements édités tandis qu’un léger son d’aérodynamique est synthétisé et ajouté au mixage de l’ensemble. Des effets (filtres, *pitchshifters*) appliqués sur les sons de roulements et d’aérodynamiques modifient le rendu en fonction du

type de véhicule ainsi que de sa vitesse de déplacement. La fenêtre d'interface *Engine Designer* de Fmod (Fig.5.4) permet de tester différents réglages afin d'obtenir le rendu le plus lisse et réaliste possible. Les événements sonores (klaxon et/ou avertisseurs, freinage brusque) sont enregistrés et édités séparément, certains bénéficient de variations de hauteur aléatoires afin d'ajouter de la diversité dans l'environnement, d'autres (klaxon et cloche de bus, sirènes) restent naturels afin de préserver leur similitude avec les sons réels.



FIGURE 5.4 – Fmod : fenêtre *Engine Designer*.

Chacune de ces composantes possède ses propres paramètres de diffusion 3D afin d'obtenir le rendu le plus crédible possible. Ces différentes strates sonores prennent en entrée les variables envoyées par le noyau de la simulation de Terra Dynamica à chaque trame (nombre de tours/minutes, vitesse du véhicule et type de surface sur laquelle se trouve le véhicule) ainsi que le déclenchement des événements (klaxon, freinages...).

Sélection dynamique des sources

Cependant, pour des raisons techniques et de temps de calcul, toutes les sources de la scène ne sont pas actives simultanément. La limitation en polyphonie de Fmod, c'est-à-dire en nombre de sources qu'il est possible

de sonoriser simultanément, demande de choisir quelles sources sont prioritaires pour la sonorisation en cas de surnombre. Un système de priorité générique prend en compte la proximité de la source au point d'écoute ainsi qu'un facteur de poids attribué à chaque type de son. Il est ainsi possible de maintenir audibles des sources très lointaines mais pertinentes pour la scène (comme une sirène) tout en rendant silencieuses des sources considérées moins importantes ou masquées (comme certains piétons par rapport aux véhicules).

En raison de la densité que peuvent atteindre les scènes de Terra Dynamica, ce système de priorité est complété d'une frontière virtuelle articulant le mixage entre Fmod et le synthétiseur d'arrière-plans. Hormis quelques exceptions, seules les sources se trouvant dans un certain rayon autour du point d'écoute sont prises en compte par Fmod, les autres étant ignorées car considérées comme trop lointaines pour être clairement entendues ou masquées par les sources voisines. Cette sélection est réalisée dans l'interface bas-niveau de Fmod, c'est-à-dire la partie codée⁴ servant à faire le lien entre Terra Dynamica et l'interface haut-niveau de Fmod.



FIGURE 5.5 – Terra Dynamica : rendu visuel (1/2)

Fmod sonorise les sources individuelles dans un rayon d'environ 100 mètres défini autour de la caméra. Les sources présentes dans la scène au-delà de ce périmètre ne sont pas prises en compte par Fmod mais font partie des données de description de la scène transmises au synthétiseur, ce dernier en tient compte pour gérer la densité de

4. Cette partie est réalisée techniquement par Shih-Han Chan [CNTT12] et Guillaume Levieux.

l'arrière-plan (voir paragraphes suivants). Au-delà de cette frontière, les sources sont considérées comme incluses dans le rendu du synthétiseur. Certains types d'événements sonores échappent cependant à ce filtrage en raison de leur nature, tels que les klaxons ou certains avertisseurs (sirènes) dont les distances maximales de diffusion sont définies dans Fmod.

Cette stratégie opère donc une distinction entre deux manières de décrire une même scène sonore. D'un côté, les sources individuelles directement observables qui constituent les actions de la scène. D'un autre côté, la synthèse qui repose sur des paramètres décrivant l'ensemble de la scène et donc son état à un instant donné.

5.1.2.2 Intégration du synthétiseur

Le rôle du synthétiseur d'ambiances urbaines est de créer un support pour les sources gérées par Fmod en rendant la densité perceptive de la scène. Il génère un arrière-plan complémentaire du flux sonore produit par Fmod. Fmod fournit des outils de synthèse granulaire mais ceux-ci ne permettaient pas de parvenir au même résultat que le synthétiseur que nous proposons, notamment en termes de mixage dynamique et de flexibilité dans les paramètres de granulation.

L'intégration du synthétiseur se fait en deux étapes principales. La première consiste à choisir les fichiers audio utilisés par le synthétiseur. La seconde étape vise à lier les paramètres du synthétiseur à des variables de la simulation. Nous présentons ces deux étapes dans les paragraphes suivants.

Les fichiers audio choisis pour alimenter la synthèse granulaire sont enregistrés avec la même méthode que celle décrite en 4.6. Tous ces enregistrements sont réalisés dans Paris afin de conserver la couleur sonore de la ville dans la simulation. Les fichiers choisis sont organisés en trois séries de quatre échantillons thématiques, chaque série étant attribuée à un oscillateur granulaire.

La première série est composée d'enregistrements réalisés aux abords du périphérique. Les quatre fichiers correspondent à des prises graduellement plus proches des voies de circulation. Ces enregistrements restituent une sensation de circulation et de densité. Les fichiers sont ensuite ordonnés dans l'oscillateur granulaire afin de donner une sensation de *crescendo* ou, à l'inverse, de *decrescendo* dans la densité perçue

du trafic. L'utilisation des filtres peut ensuite accentuer cette sensation jusqu'à ne restituer du flux de circulation qu'un drone très tenu et lointain.



FIGURE 5.6 – Terra Dynamica : rendu visuel (2/2)

La deuxième série est composée d'enregistrements réalisés sur la place de la République. Ces enregistrements sont effectués à différents endroits de la place, à des intensités de circulation différentes et tentent de capturer la sensation d'espace et de circulation du lieu.

Enfin, la troisième série est composée d'enregistrements réalisés aux abords de la Bibliothèque François Mitterrand. Il s'agit d'ambiances relativement ténues et peu chargées en trafic. Ce sont davantage des « fonds d'air » que des enregistrements de l'activité urbaine directe et ils peuvent apporter une coloration intéressante en cela.

Le mélange de ces trois groupes de matières différentes dans leurs quatre variations sert à obtenir une palette de textures complémentaires. Les enregistrements du périphériques donnent une base compacte qui peut servir de support aux enregistrements plus animés de la place de la République. Les enregistrements aux abords de la Bibliothèque peuvent apporter une ouverture dans les plus hautes fréquences. Dans Terra Dynamica, la caméra suit le point d'écoute et celle-ci peut se déplacer sans contraintes, il est donc possible de se trouver au ras du sol, au milieu du trafic ou virtuellement à des dizaines de mètres au-dessus de la scène. Ce facteur de *zoom* vertical peut être finement suivi avec des textures couvrant des registres allant du plus dense au tenu. Le déplacement horizontal de la caméra (et donc du point d'écoute) permet de naviguer entre des zones plus ou moins densément peuplées. Ces deux axes de déplacement se combinent et sont pris en compte

par le synthétiseur afin de mixer et traiter en temps réel les différents enregistrements.

Le mixage dynamique et la synthèse temps-réel requièrent donc une communication entre le synthétiseur et Terra Dynamica afin d'obtenir une description de la scène utilisable par le synthétiseur. Cette communication est assurée *via* un objet Pure Data conçu pour le projet et développé en collaboration avec Guillaume Levieux. Cet objet, baptisé [Terradynanimac~], reçoit un certain nombre de variables depuis Terra Dynamica et les transmet aux différents paramètres de synthèse. La communication entre Terra Dynamica et Pure Data est assurée par un protocole UDP transmettant les variables requises depuis leurs ports attribués dans la simulation vers l'objet [Terradynanimac~].

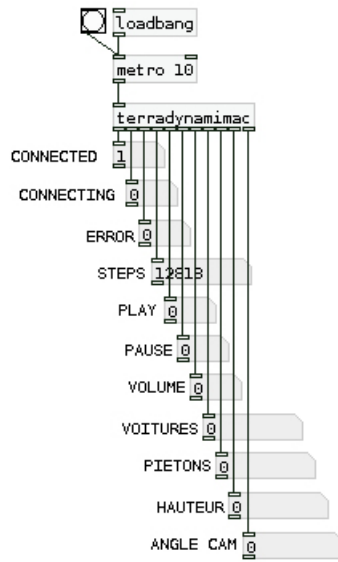


FIGURE 5.7 – Terradynanimac~ : détail des sorties de l'objet.

La Fig.5.7 montre un exemple de *patch* Pure Data avec l'objet [Terradynanimac~], les onze sorties présentes sont reliées à des boîtes de type *nombre* annotées du type de variable fournie. De gauche à droite, les deux premières renseignent sur l'état de la connexion entre la simulation et l'objet, la troisième renseignant sur une éventuelle erreur dans la communication. La quatrième sortie donne le nombre de

pas de simulation traités : l'objet `[metro]` permet de régler la fréquence à laquelle les informations sont appelées depuis la simulation vers Pure Data ; dans l'exemple de la Fig.5.7, `[Terradynanimac~]` transmet les données reçues de la simulation par intervalles de 10 millisecondes. Les cinquième et sixième sorties (annotées *Play* et *Pause*) renseignent l'état de la simulation. La septième sortie fournit une information générique de volume entre 0 et 1. Les deux sorties suivantes donnent le nombre de véhicules et de piétons présents dans une sphère d'environ 200 mètres autour du point d'écoute. Les sorties dix et onze renseignent respectivement la hauteur de la caméra par rapport au sol (en mètres) et l'azimut de la caméra.

Les sorties du `[Terradynanimac~]` sont reliées à différents paramètres du synthétiseur. Les sorties relatives à l'état de la simulation (annotées *Play* et *Pause*) servent à la mise en route et à l'arrêt du synthétiseur. Les sorties *voitures* et *piétons* sont utilisées pour contrôler des paramètres de mixage à l'intérieur de chaque oscillateur granulaire ainsi qu'entre les oscillateurs. La sortie *hauteur* est utilisée pour modifier le mixage entre les trois oscillateurs granulaires ainsi que les paramètres de filtrage sur chacun des oscillateurs.

La variation des paramètres en temps réel peut connaître des sauts de valeur assez importants. Le lissage des valeurs est toujours possible mais seulement dans la mesure où il ne crée pas un décalage entre le rendu sonore et le rendu visuel. Il est donc apparu essentiel que les paramètres du synthétiseur résistent à des variations brutales de valeur sans créer d'artefacts audibles. Le temps de lissage choisi varie entre 50 et 100 millisecondes (selon le paramètre impacté), ce qui laisse suffisamment de temps au synthétiseur pour effectuer des transitions fluides entre des valeurs extrêmes sans que cela ne semble nuire à la crédibilité perceptive de l'arrière-plan sonore.

Une fois lissées, les variables ne sont pas utilisées brutes mais le plus souvent normalisées pour le besoin du synthétiseur. La normalisation permet également de fixer des minima et maxima arbitraires ainsi que d'influencer la linéarité des valeurs entre les deux extrêmes. L'étape de normalisation est faite sur la base d'ajustements empiriques prenant en compte la crédibilité perceptive des masses sonores générées, leur réactivité et leur mélange avec les sources sonorisées par Fmod.

5.1.3 Bilan et conclusion

L'objectif dans la sonorisation de Terra Dynamica était de compléter les fonctionnalités de sonorisation en 3D de Fmod par un outil de synthèse permettant de générer des arrière-plans sonores d'ambiances urbaines. L'intégration des deux solutions en parallèle s'est révélée fonctionnelle et réactive dans le cadre du démonstrateur. Le démonstrateur comprenait le 5^{ème} arrondissement de Paris ainsi que la Place de la République, soit des zones avec des densité de trafic et des espaces très divers.

Parmi les améliorations notables, il serait intéressant de pouvoir créer des groupes de tailles intermédiaires entre les sources individuelles et l'arrière-plan. Les groupes permettrait peut-être d'accentuer l'effet de profondeur de champ mais aussi de caractériser des groupes d'acteurs clairement identifiés (par exemple des groupes de manifestants). La principale difficulté à la réalisation de ces fonctionnalités concernant les conditions d'agrégation et de désagrégation des groupes d'individus ou de véhicules, régies par le noyau d'intelligence artificielle de Terra Dynamica.

L'autre point qui mériterait sûrement des développements plus poussés concerne les aspects acoustiques de la simulation. Afin de placer les sources individuelles dans un espace acoustique proche de celui d'une ville, nous avons principalement utilisé les réverbérations algorithmiques fournies avec Fmod. Si le résultat obtenu avec cette solution est convainquant, il n'est pas aussi exact que souhaité.

Enfin, pour accroître la finesse du rendu proposée par la synthèse, il serait nécessaire de prendre en compte davantage d'aspects sémantiques dans la description de scène fournie par le noyau de la simulation. La proportion de chaque type de sources présentes ainsi que la nature de l'environnement (grand espace ouvert ou ruelle) devraient avoir un impact plus marqué sur la synthèse des arrière-plans.

La sonorisation de Terra Dynamica relevait d'un partage entre un environnement virtuel proche de la simulation et des technologies de réalisation fortement inspirées de celles du jeu vidéo. Si des compromis furent nécessaire entre ces deux aspects, la sonorisation réalisée semble être suffisamment efficace et a pu mettre en avant les fonctionnalités du synthétiseur d'ambiances dans un environnement interactif.

5.2 Écoutes comparatives

La seconde étape d'expérimentation sur le synthétiseur consiste à réaliser un test perceptif. Ce test se situe dans la continuité des méthodologies utilisées pour la validation des méthodes de synthèse de textures sonores, telle que nous avons pu les référencer précédemment (voir 2.5). Cette étape vise à appréhender la qualité de la re-synthèse proposée en comparant son rendu à celui produit par d'autres algorithmes.

Les études générales sur la perception des textures sonores environnementales re-synthétisées sont rares [Sch11], elles incluent cependant de plus en plus souvent des évaluations perceptives relatives à la qualité de la synthèse sonore. Dans leur majorité, les outils de synthèse de texture développés sont pensés relativement à un type de texture prédéfini. Les protocoles expérimentaux visant à valider la synthèse de texture reposent donc principalement sur l'écoute comparative entre les enregistrements d'origine et leur re-synthèse [Mas11].

Nous supposons que notre méthode de synthèse de textures peut s'appliquer à une grande variété de sources sonores mais l'expérimentation que nous décrivons dans cette section se focalise sur l'objectif premier du synthétiseur, c'est-à-dire les ambiances urbaines. Nous proposons donc un test d'écoute comparative de fichiers sonores produits par le synthétiseur avec d'autres méthodes de re-synthèse, sur la base de fichiers de départ identiques.

Pour cette expérience, l'évaluation du rendu du synthétiseur dans son ensemble nous a paru constituer une tâche complexe en raison de sa multitimbralité. En effet, il peut y avoir jusqu'à douze lecteurs granulaires actifs en simultané et la masse sonore ainsi générée pourrait rendre indiscernable ce qui relève du synthétiseur et ce qui relève d'un effet produit par la superposition des timbres. D'autre part, si cet effet de superposition était volontairement recherché dans le design du synthétiseur, ce même design fait que, s'il est possible d'entendre jusqu'à douze lecteur granulaires superposés, il est également possible, à l'opposé, d'en entendre un seul.

Il nous a donc semblé plus pertinent, dans un premier temps, d'évaluer le lecteur granulaire seul (*Grainer*) plutôt que le synthétiseur complet. Nous émettons l'hypothèse que l'objet *Grainer* constitue un élé-

ment autonome répondant aux critères d'un générateur de textures sonores.

Dans les chapitres précédents nous avons vu que l'un des principes de la synthèse de texture est de segmenter un signal de départ puis de composer, en réorganisant les différents segments sur une durée indéterminée, un signal perceptivement similaire à l'original. L'objet *Grainer* propose une méthode pour parvenir à ce résultat en sélectionnant aléatoirement et en mélangeant des grains extraits d'un fichier de départ. Cependant, la re-synthèse produite par cette méthode est susceptible de dégrader la qualité du fichier d'origine à deux niveaux. Premièrement, en segmentant le fichier d'origines en grains, l'algorithme risque d'induire des effets de rupture et discontinuité rendant le résultat artificiel ou incohérent. Pour pallier cet effet, la re-synthèse fait se superposer les grains afin de lisser leur enchaînement. Cette superposition est potentiellement la deuxième cause de dégradation dans la mesure où elle induit des effets de mixages qui ne sont évidemment pas présents dans le fichier d'origine.

Le design de l'objet *Grainer* prend en compte la nature des sons utilisés par le synthétiseur. Ainsi, les paramètres de re-synthèse ont fait l'objet de réglages afin de minimiser l'impact perceptif des transformations induites. L'objectif de cette expérimentation est donc de valider la méthode de synthèse proposée en vérifiant qu'elle ne dénature pas significativement un échantillon de départ au cours de son processus.

Dans cette section nous détaillerons une première expérimentation sur le rendu sonore de l'objet *Grainer* en présentant le protocole et les passations, les résultats obtenus et enfin une discussion portant sur l'expérimentation dans son ensemble.

5.2.1 Expérimentation

Le protocole choisit pour évaluer l'efficacité de notre méthode de re-synthèse repose sur des écoutes comparatives entre des enregistrements et leur version re-synthétisée. Cependant, la comparaison perceptive entre un échantillon et une re-synthèse seule risque de ne pas suffire à renseigner sur les aspects de la re-synthèse qui s'avèrent positifs ou négatifs. La comparaison entre un échantillon et plusieurs re-synthèses de ce même échantillon (telle que présentée dans [Mas11] p.10) semble une stratégie intéressante pour une évaluation plus fine.

Nous proposons donc que le test repose sur la comparaison perceptive entre un enregistrement et différentes re-synthèses de celui-ci. Ces re-synthèses incluent notre méthode et des re-synthèses utilisant une méthode suffisamment proche. Pour la réalisation de ces dernières, le choix s'est porté sur *timbreID* [Bre10].

TimbreID est une librairie développée pour Pure Data permettant d'aborder la synthèse concaténative. Cette méthode de synthèse, dérivée de la synthèse granulaire, permet d'obtenir des résultats convainquant pour la re-synthèse de textures sonores [Sch11]. Utiliser *timbreID* est également un moyen de comparer deux solutions dans le même environnement : Pure Data.

Dans un premier temps nous détaillons le choix des échantillons utilisés pour l'expérimentation (enregistrement et édition). Dans un second temps, nous abordons plus en détails le protocole et décrivons les conditions de passation de l'expérimentation.

5.2.1.1 Choix des échantillons pour l'expérimentation

Dans la continuité du développement du synthétiseur, le choix des matériaux sonores utilisés pour l'expérimentation se porte sur des ambiances urbaines. Ces ambiances correspondent au plus près à la notion d'arrière-plan, il s'agit de textures uniformes et diffuses desquelles aucune source unique et clairement identifiable n'émerge. Les échantillons sont choisis dans la bibliothèque d'enregistrements urbains constituée au cours de la thèse. Cette bibliothèque est réalisée à l'aide d'un enregistreur portable Olympus LS-5 dont les réglages sont identiques pour chaque enregistrement⁵. Les lieux d'enregistrement sont choisis en fonction du paysage sonore qu'ils proposent à entendre, il s'agit le plus souvent de grands espaces ouverts ou situés en hauteur et éloignés de sources sonores individuelles.

Quatre échantillons sont sélectionnés pour l'expérimentation. Les critères de sélection sont, pour chaque échantillon, l'uniformité de la dynamique et la stabilité du timbre d'une part et, d'autre part, la constitution d'une palette de quatre couleurs sonores différentes. Une fois choisis, les échantillons sont ramenés à une durée de 20 secondes. Le choix de cette durée repose sur deux contraintes. La première contrainte

5. Enregistrement au format *.wav* dans une résolution de 48 kHz et 24 bits, niveau des pré-amplificateurs réglé à 4 sur 10 (sensibilité «*low*») et désactivation du filtre passe-haut.

est que le fichier doit être suffisamment long pour obtenir un nombre significatif de grains pour la re-synthèse. La seconde contrainte est que les fichiers doivent être d'une durée suffisamment courte pour l'expérimentation : des fichiers plus longs risquant d'allonger la durée des passations et donc de causer une certaine fatigue chez les écoutants.

Les quatre échantillons sélectionnés (nommés ci-après A, B, C et D) sont re-synthétisés avec deux méthodes différentes afin de constituer un corpus total de vingt fichiers sonores : quatre échantillons originaux et quatre resynthèses pour chacun d'entre eux. Trois de ces re-synthèses sont réalisées à l'aide de la librairie *timbreID*, selon trois réglages différents afin d'obtenir des variations de re-synthèse (principalement l'organisation des grains et leur ordre de lecture). La quatrième re-synthèse est produite à l'aide du *patch Grainer* (voir 4.4.2). L'Annexe C détaille la réalisation de ces échantillons avec les deux méthodes.

Une fois les échantillons produits, ceux-ci sont organisés dans une interface Pure Data créée pour l'expérimentation et permettant leur lecture selon les conditions souhaitées. Cette interface est décrite dans les paragraphes suivants, ainsi que le questionnaire accompagnant les écoutes.

5.2.1.2 Protocole

L'expérimentation est composée de deux éléments : une série de *patches* Pure Data (servant d'interface pour l'écoute des fichiers sonores) et un questionnaire sur papier à remplir par l'écouter (voir Annexe D). Les passations sont réalisées dans les locaux de l'Enjmin⁶, dans une salle isolée. L'écoute est faite au casque (modèle AKG K55) sur un ordinateur portable. L'expérimentation cumule un total de 24 participants, tous étudiants en Master à l'Enjmin, âgés de 19 à 30 ans (avec une moyenne d'âge de 24 ans) et ne témoignant pas de trouble auditif.

Le test proposé aux participants est de procéder à quatre écoutes comparatives entre un son de référence et quatre autres sons. L'écouter n'est pas informé de la relation entre les sons de référence et les autres sons proposés.

La consigne donnée est d'évaluer, pour chacun des quatre sons à comparer, la similarité avec le son de référence. Pour cela, l'écouter reporte une note associée à un son sur une échelle de Likert allant de

6. Ecole Nationale du Jeu et des Médias Interactifs, à Angoulême

1 (« Pas du tout proche ») à 10 (« Extrêmement proche »). Cette tâche est répétée avec les quatre sons de référence, ce sont donc quatre séries de sons qui sont présentées à l'écouter, référencées A, B, C et D, conformément au choix des échantillons détaillé dans le paragraphe précédent.

Le son de référence de chaque série peut être écouté autant de fois que souhaité et les quatre sons à comparer peuvent être écoutés trois fois chacun. L'écouter est d'abord invité à régler le volume d'écoute selon sa convenance. Puis il est encouragé à prendre le temps d'écouter l'intégralité des fichiers, dans l'ordre qu'il souhaite. Il est également précisé qu'il n'y a ni bonne ni mauvaise réponse.

Afin de ne pas présenter les fichiers sonores dans le même ordre pour tous les participants, les séries A, B, C et D sont déclinées dans quatre versions, chacune proposant une disposition différente des re-synthèses. Les versions sont alternées tous les trois participants tout au long des 24 passations réalisées.

Chacune des versions est donc utilisée six fois au total. Si l'ordre des sons à l'intérieur de chaque série variait, l'ordre des séries est par contre resté le même (A, puis B, puis C et enfin D).

La Figure 5.8 est une capture d'écran de l'interface de test réalisée dans Pure Data et proposée aux participants pour entendre les fichiers audio. Le son d'origine est sur la partie gauche de la fenêtre, les fichiers de re-synthèse sont sur la partie droite. Une glissière horizontale représente le temps de lecture de chaque fichier. Sur la partie droite, une boîte numérique affiche le nombre restant de lectures possibles pour chacune des re-synthèses. La mise en lecture d'un fichier audio se fait en cliquant dans l'une des boîtes carrées renseignées « Play/Stop », cliquer dans une autre boîte arrête automatiquement la lecture du fichier en cours et déclenche celle du fichier sélectionné.

Pour chaque écouter, la passation durait en moyenne entre quinze et vingt minutes, avec généralement davantage de temps consacré à la première série qu'aux trois suivantes.

5.2.2 Présentation et analyse des résultats

Le tableau présentant le détail des résultats sous la forme des scores attribués par chaque écouter à chaque re-synthèse est en Annexe E.

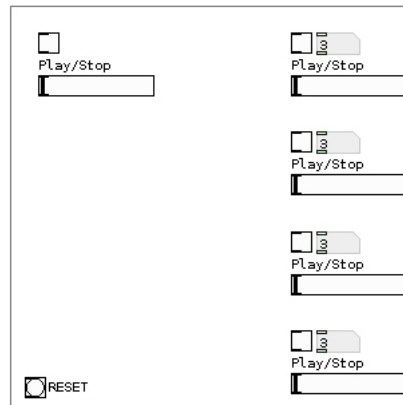


FIGURE 5.8 – Pure Data : *patch* utilisé comme interface de test pour les passations

Le numéro des participants est indiqué dans la colonne de gauche. Le groupe de colonnes A1 à A4 correspond aux notes attribuées respectivement aux re-synthèses *centroid/amplitude* (A1), *centroid/brightness* (A2), *centroid/flatness* (A3) et *Grainer* (A4) pour l'échantillon A. Il en va ensuite de même pour les re-synthèses des échantillons B, C et D.

Nous proposons de traiter cette première présentation en calculant les moyennes et écart-types pour chaque colonne afin de comparer les résultats de chaque échantillon.

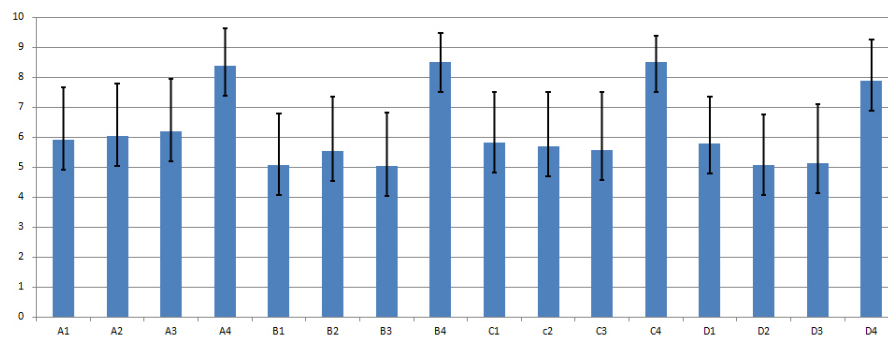


FIGURE 5.9 – Moyennes et écart-types des résultats de l'expérimentation.

La Figure 5.9 montre les moyennes et les écart-types des notes

obtenues pour chacun des sons A, B, C et D dans leurs quatre variantes de re-synthèses annotées 1, 2 et 3 pour *timbreID* et 4 pour *Grainer*.

Les résultats obtenus par les trois variantes de *timbreID* semblent relativement uniformes pour chacun des quatre échantillons. Les résultats obtenus par *Grainer* semblent également uniformes et globalement supérieurs à ceux obtenus par *timbreID*.

Nous complétons ces observations par un test d'hypothèse statistique afin de confirmer ou d'infirmer que l'écart observé entre les moyennes de notre méthode et celles de *timbreID* est significatif. Le recours à ce test est justifié notamment par les écart-types importants observés sur les notations de *timbreID*. Pour chaque enregistrement, nous comparons la moyenne attribuée à notre méthode et la moyenne attribuée à chaque échantillon *timbreID* en posant deux hypothèse :

- hypothèse nulle : la moyenne attribuée à notre méthode est équivalente à celle obtenue par *timbreID*
- hypothèse alternative : la moyenne attribuée à notre méthode est différente de celle obtenue par *timbreID*

L'objectif de ce test est donc de comparer deux moyennes provenant de petits *échantillons* (inférieurs à 30) dont la variance de la *population* n'est pas connue. Nous choisissons donc d'utiliser le test paramétrique suivant la loi *t* de Student, dérivée de la loi Normale.

Après le calcul de la variance pour les seize séries de résultats, nous observons des écarts importants entre notre méthode et *timbreID*. Le test de Student que nous utilisons incorpore donc une approximation de Cochran-Cox (soit une approximation du seuil de significativité) afin de prendre ces écarts en compte et considérer les variances comme potentiellement différentes.

L'Annexe F présente les résultats obtenus pour les douze tests effectués. Ces résultats nous permettent d'affirmer que l'hypothèse nulle peut être rejetée au seuil 0,05 (avec $p < 0.0001$) au profit de l'hypothèse alternative, les moyennes obtenues par notre méthodes semblent donc significativement supérieures à celles obtenues par *timbreID*. Cette affirmation doit cependant être nuancée et discutée.

5.2.3 Discussion et conclusion

À la première observation de ces résultats, on note une distinction entre les évaluations accordées à *timbreID* et celles accordées à *Grainer*. Il est également intéressant de noter l'uniformité des moyennes des trois re-synthèses réalisées avec *timbreID*, les variations dans la réalisation de ces re-synthèses ne semblant pas avoir eu un impact significatif sur la qualité perçue. Si ces résultats -montrent au premier abord un avantage en faveur de la méthode que nous proposons, leur interprétation doit être pondérée. En effet, les meilleures notations attribuées aux fichiers réalisés avec *Grainer* ne signifient pas que cette méthode est dans l'absolu meilleure que le système de *timbreID*. Rappelons également que le système de re-synthèse proposé dans le patch « *timbre-space* » (utilisant *timbreID*) n'est pas calibré spécialement pour re-synthétiser un type de son en particulier mais davantage pour être adaptable à la palette de sons la plus vaste possible. Les moyennes seules ne suffisent donc sans doute pas à rendre compte objectivement des évaluations.

Les écart-types sont systématiquement plus élevés sur les moyennes des sons produits avec *timbreID*, l'observation des notations individuelles permet d'y apporter une éventuelle interprétation. S'il est rare que les sons produits par *Grainer* soient moins bien notés que ceux produits avec *timbreID*, il est en revanche fréquent de trouver des notations très proches ou égales entre les deux méthodes. L'écart observé entre les moyennes ne se retrouve pas chez tous les écoutants et tous ne notent donc pas de différence significative entre les rendus sonores des deux méthodes.

Si la consigne était de noter les re-synthèses individuellement et relativement au fichier d'origine, il ne nous est pas possible de dire si cette consigne était intégralement respectée. Telle que l'expérience est menée, nous ne savons pas si les écoutants ont noté les différents fichiers de re-synthèse relativement au fichier d'origine ou relativement entre eux. C'est-à-dire qu'un écoutant peut attribuer la note maximale à une re-synthèse (parce qu'elle lui semble être la meilleure des quatre, sans pourtant qu'elle soit équivalente au fichier d'origine) et noter ensuite les autres re-synthèses relativement à ce choix. Lorsque interrogés sur leur ressenti après la passation, certains des participants ont effectivement exprimé l'impression de « noter les quatre sons les uns par rapport aux autres. »

Ces résultats tendent cependant à montrer la relative efficacité de la méthode de synthèse granulaire développée au cours de cette thèse, dans le cas de la re-synthèse des textures choisies pour l'expérimentation. Les notations sont relativement uniformément positives et semblent appuyer l'efficacité des paramètres de re-synthèse choisis.

Il est également intéressant de noter que la sélection aléatoire des grains lors de la re-synthèse ne semble pas poser de problèmes qualitatifs majeurs. Ceci s'explique sans doute en partie par le travail d'édition effectué sur les enregistrements avant de les utiliser comme sources pour le synthétiseur. Le temps de recouvrement relativement long entre les grains (ici il s'agit de la taille du grain divisée par 2.571) peut expliquer en partie cette efficacité.

Dans la perspective d'affiner ces premiers résultats, des consignes différentes ou un changement de protocoles devraient être nécessaires afin de contrôler plus finement l'évaluation donnée par les écoutants, c'est à dire la similarité entre le son d'origine et une re-synthèse et non les re-synthèses entre elles.

D'autre part, si cette expérimentation semble confirmer que la méthode que nous proposons pour re-synthétiser des textures sonores ne dégrade pas perceptivement le fichier d'origine, elle ne donne pas assez d'indices sur l'optimisation des réglages des deux principaux paramètres, à savoir la taille des grains et le temps de recouvrement entre les grains. De nouvelles expérimentations devraient donc être menées afin de déterminer de quelle manière ces deux paramètres influencent le rendu perceptif de la re-synthèse, s'il existe une palette de réglages optimale et si cette palette varie d'un son à un autre.

Il convient de rappeler également que ces évaluations portent sur la composante algorithmique de l'élément *Grainer* et non sur l'ensemble du synthétiseur. Ce test rend donc partiellement compte du composant le plus bas-niveau du synthétiseur et d'autres tests seraient nécessaires pour une approche critique du rendu multitimbral du synthétiseur. Le test le plus complet devrait être effectué dans le cadre d'une ville virtuelle interactive. Cependant, un tel protocole peut s'avérer complexe à élaborer, notamment en raison du manque de références sur l'évaluation perceptive de textures sonores dynamiques dans le cadre d'applications vidéoludiques.

L'intégration et les écoutes comparatives devraient également être

complétés par des tests d'ordre plus techniques sur la résistance de la méthode de re-synthèse que nous proposons. L'intégration du synthétiseur à Terra Dynamica n'a apparemment pas induit de problèmes de performance. Nous avons régulièrement vérifié, au cours de cette étape, que les ressources utilisées par le synthétiseur restaient raisonnable (environ trois pour cent).

Nous n'avons cependant pas effectué de tests de passage à l'échelle outre mesure car, dans l'immédiat, ces constatations semblaient suffisantes. Elles ne suffisent cependant pas à rendre compte de la capacité maximale d'exploitation du synthétiseur et des optimisations qui seraient nécessaires dans le cas d'une exploitation plus intensive. Des tests de passage à l'échelle seraient nécessaires pour estimer, par exemple, le nombre de *Grainers* potentiellement utilisables en simultané dans une scène ou, plus généralement, la capacité de sonorisation totale qu'il est possible d'atteindre avec le synthétiseur. Il apparaît donc indispensable de conduire de tels tests dans de futurs travaux.

Enfin, nous avons envisagé de recourir à des supports visuels pour le test passé. Cependant, la présence d'un tel support risquait d'induire d'autres biais que nous n'étions pas certains de contrôler et la décision était prise de faire un test uniquement audio. Il apparaîtrait pourtant logique, à terme, de tester la pertinence du rendu sonore du synthétiseur sur des visuels, notamment parce que les deux perceptions (sonore et visuelle) semblent complémentaires pour l'appréciation et l'appréhension de l'environnement urbain, réel ou virtuel ([Thi02], [KM10], [KHP⁺13]).

Ce serait ici une ouverture vers une tentative de mesure qualitative du facteur d'immersion proposé par le synthétiseur, ce facteur étant important dans les expériences vidéoludiques ([Gri08], [Tho10]).

Chapitre 6

Conclusion générale et perspectives

Sommaire

6.1 Contributions	181
6.1.1 L'arrière-plan comme objet sonore	181
6.1.2 Réalisation d'un synthétiseur d'arrière-plans sonores urbains dynamiques	182
6.2 Analyse critique et travaux futurs	183
6.2.1 Développements du synthétiseur	183
6.2.2 Validations expérimentales	186
6.2.3 Intégrations futures	187

Nous avons vu que la création des arrière-plans sonores dans les médias interactifs reposent principalement sur des techniques de mixage proches de celles utilisés dans les médias audio-visuels linéaires. Cependant, l'utilisation de techniques procédurales permettant la génération de sons selon les besoins de la scène semble de plus en plus à même de répondre aux besoins des environnements virtuels complexes tant dans leur contenu que dans leur interactivité.

Le principal objectif de ce travail de thèse était donc d'explorer une méthode, une architecture et des stratégies de sonification temps-réel pour la génération d'arrière-plans plans d'ambiances sonores urbaines. Dans ce dernier chapitre nous abordons les contributions apportées par ce travail de thèse ainsi que les travaux futurs envisagés.

6.1 Contributions

À l'issue de ce travail de thèse, nous proposons de retenir deux contributions principales. Dans un premier temps, il nous a été possible de mieux cerner le rôle des arrière-plans sonores dans l'écriture de scènes urbaines virtuelles. Les arrière-plans constituent des objets sonores à part entière, parfois complexes, et leur rôle dans la narration et les informations fournies les distinguent des autres objets sonores.

Ensuite, nous avons proposé une architecture de synthèse granulaire tournée spécifiquement vers les textures sonores d'arrière-plan. Ce prototype a pu être implémenté et testé afin de valider, au moins partiellement, ses fonctionnalités.

6.1.1 L'arrière-plan comme objet sonore

Alors que la notion de plan sonore est totalement intégrée dans l'audiovisuel linéaire nous avons montré qu'elle est tout aussi pertinente dans l'écriture sonore de villes virtuelles interactives telles que celles rencontrées dans les jeux vidéo.

Nous avons vu que dans la sonorisation de scène urbaines virtuelles, l'arrière-plan sonore devient un objet à part entière. Son rôle est d'apporter une couleur à l'ambiance de la scène tout en servant de liant aux autres sons présents, il témoigne d'un ressenti complétant ce qui est visible. Il permet aussi de restituer une description sonore de l'état

de la scène, ce qui le différencie des sources individuelles, ces dernières décrivant principalement des actions ponctuelles.

L'arrière-plan a une place dans l'écriture sonore des environnements interactifs, il peut être intégré aux stratégies de sonification d'environnements urbains interactifs complexes. La création d'un arrière-plan dynamique constitue un processus d'écriture sonore distinct de ceux des autres sons d'une scène, par son ubiquité, sa capacité à apporter des informations que le visuel ne peut restituer ainsi qu'à créer de l'immersion en prolongeant la ligne d'horizon sonore.

Le processus de création des arrière-plans sonores dynamiques implique le plus souvent la conception d'outils spécifiques dont les paramètres sont pensés relativement aux fonctions des arrière-plans eux-mêmes dans l'environnement.

6.1.2 Réalisation d'un synthétiseur d'arrière-plans sonores urbains dynamiques

En considérant les propriétés des arrière-plans sonores urbains, nous avons conçu un prototype de synthétiseur granulaire dynamique permettant de reconstituer des ambiances sur la base d'enregistrements.

Le principal objectif était de donner au designer sonore les moyens de gérer dynamiquement le contenu et la présence de l'arrière plan en tenant compte des paramètres dynamiques fournis en temps réel, par le moteur de jeu, ainsi que des déplacements du point d'écoute dans un système interactif.

La méthode de granulation est construite autour des observations et analyses des mécanismes et de la morphologie sonores des arrière-plans sonores urbains. Nous avons ainsi privilégié l'utilisation de grains relativement et d'un mixage dynamique de plusieurs *Grainers*. Il ne s'agit pour autant pas d'une solution fermée : l'architecture est suffisamment souple pour répondre à différentes configurations de création et d'intégration d'arrière-plans sonores.

Nous avons également détaillé la méthodologie d'enregistrement utilisée pour alimenter le synthétiseur en textures à re-synthétiser. Il n'existe, à notre connaissance, pas de méthodologie décrivant scientifiquement l'enregistrement d'arrière-plans sonores urbains. Cette méthodologie est donc empirique, elle repose autant sur notre propre expérience que sur les descriptions proposées par les designers sonores, avec leur sensibilité propre. L'architecture du synthétiseur vise à créer

un lien entre des besoins technologiques et des pratiques spécifiques à chaque designer sonore.

Nous avons ensuite pu expérimenter le synthétiseur dans le contexte du projet Terra Dynamica. En complément des fonctionnalités du moteur audio Fmod, le synthétiseur a grandement participé au contenu des paysages sonores dynamiques. L'ajout de textures urbaines diffuses re-synthétisées en temps réel a consolidé le rendu fourni par Fmod tout en accentuant la sensation de perspective acoustique ou encore en soulignant les effets d'élévation de la caméra par rapport au sol.

D'autre part, les premiers tests perceptifs effectués sur le rendu du synthétiseur, en comparaison d'enregistrements réels et d'une autre technique de re-synthèse, tendent à montrer la relative efficacité de la méthode proposée.

6.2 Analyse critique et travaux futurs

Les contributions que nous venons de présenter ne constituent cependant que des étapes préliminaires qui mériteraient de plus amples développements.

Dans la perspective de poursuivre les travaux entamés au cours de cette thèse, nous présentons ici de futurs axes de développement possibles pour le synthétiseur, le besoin de pousser les expérimentations perceptives et enfin les prochaines intégrations du synthétiseur.

6.2.1 Développements du synthétiseur

Le synthétiseur réalisé couvre une partie des objectifs de ce travail de thèse. Il reste cependant améliorable et un certain nombre de points qui mériteraient d'être développés dans de futurs travaux afin d'accroître la qualité du rendu et les possibilités de design sonore. D'autre part, des développements non-intégrés au prototype pourraient être repensés afin d'enrichir la solution actuelle.

Ces développements et questionnements concernent principalement les fonctionnalités bas-niveau de la granulation, les notions de spatialisations et l'ajout de nouvelles sources sonores au synthétiseur.

6.2.1.1 Différentes options de fenêtrage des grains

Le choix du type de fenêtre dynamique appliquée aux grains est un facteur important du rendu d'une méthode de re-synthèse granulaire. L'architecture que nous avons présentée ici utilise des courbes dites de puissance égale pour produire un flux audio le plus stable possible.

Si cette logique, combinée aux autres spécifications du processus que granulation que nous avons décrites, a prévalu pendant le développement du synthétiseur, il ne serait pas exclu de tester d'autres formes de fenêtres, et même d'en combiner certaines. Nous émettons l'hypothèse que ceci permettrait, par effet de juxtaposition, de complexifier le rendu perçu du synthétiseur et d'aborder d'autres types de textures sonores. Nous menons actuellement des expérimentations dans cette perspective.

6.2.1.2 *Warping* audio

Le terme *warping* audio recouvre un ensemble de manipulations applicables à un échantillon. Il s'agit principalement de modifier la durée d'un son sans altérer sa tonalité (cette opération est le plus souvent appelée *time-stretch*) ou, à l'inverse, de modifier la hauteur tonale d'un son sans en changer la durée (*pitch-shift*). La re-synthèse granulaire est un moyen de parvenir à de tels résultats et son implémentation s'est montrée efficace dans de nombreuses réalisations commerciales comme le logiciel Live développé par la société Ableton¹

Le synthétiseur n'intègre pour le moment aucun traitement de ce type mais le travail de re-synthèse des textures urbaines pourrait gagner en variété si ces fonctionnalités étaient mises à disposition.

Il s'agirait de fonctionnalités complémentaires à la méthode de granulation précédemment décrite, visant à enrichir les possibilités de manipulation des échantillons et la diversité des ambiances produites sur cette base de données.

Une contrainte à l'implémentation de ces algorithmes concerne l'utilisation temps-réel et, si cela est possible, le temps de calcul induit. Il s'agit de perspectives de recherche fortement intéressantes pour les développements futurs du synthétiseur granulaire.

1. <https://www.ableton.com/en/live/new-in-9/> Et plus précisément sur les algorithmes de *warping* audio intégrés à Live : <https://www.ableton.com/en/manual/audio-clips-tempo-and-warping/>

6.2.1.3 Mise en espace

Les textures générées par le synthétiseur possèdent majoritairement un caractère ubiquitaire, elles ne sont pour autant pas dénuées de mouvements internes et l'aspect spatial propre à la notion d'arrière-plan mériterait de plus amples développements. Plusieurs angles d'approches possibles sont proposés ci-après.

Si le drone urbain apparaît statique, sa sonorité n'est pas la même selon, par exemple, la vitesse moyenne du trafic. Un paramètre contrôlant la quantité de mouvement dans le drone mais également la vitesse de celui-ci et sa position dans l'espace serait un ajout non négligeable à la qualité perceptive et au réalisme dynamique de la synthèse. Cela permettrait de croiser la dimension de densité avec celle de mouvement.

Implémenter une fonctionnalité permettant de contrôler dynamiquement le positionnement de chacun des oscillateurs granulaires dans l'espace de rendu serait un moyen éventuel de parvenir à ce résultat. Afin que cette méthode soit efficace, il conviendrait de pouvoir contrôler la largeur qu'occupe un oscillateur du synthétiseur dans l'espace de rendu.

Enfin, la perception de l'arrière-plan peut être colorée par l'espace environnant le point d'écoute. L'ajout de paramètres contrôlant le rendu acoustique de l'arrière-plan urbain, par le biais d'une réverbération algorithmique ou à convolution par exemple, serait un moyen de parvenir à ce résultat.

6.2.1.4 Nouvelles sources sonores

Certains sons clairement identifiables peuvent contribuer à caractériser un lieu. Murray Schafer utilise le terme *soundmarks* pour caractériser ces sons particuliers. Dans une ville, il peut s'agir de certains types de sirènes, de klaxons ou tout autre type de sons pouvant être entendus de loin. Un échantillonneur pourrait venir compléter le synthétiseur afin de gérer des fichiers sonores déclenchés ponctuellement ou aléatoirement et disposés dans l'espace selon les besoins de la scène. Ces fichiers ne seraient pas nécessairement liés à une action précise mais constitueraient des sources identifiables caractéristiques du lieu, de son état et de la sensation d'espace qui l'accompagne.

D'autre part, au cours du développement du synthétiseur, nous avons conduit un certain nombre d'essais à base de synthèse soustractive. Le principe était de traiter un bruit blanc à l'aide de différents

filtres passe-bande afin d'extraire des textures abstraites proches de celles observées sur les sonagrammes d'enregistrements urbains. L'addition de plusieurs filtres passe-bande contrôlés individuellement en amplitude permet de sculpter des textures et drones variés qui constituent des compléments intéressants aux enregistrements. Ces premières expérimentations peuvent être poursuivies et approfondies à l'aide d'analyses/re-synthèses plus poussées sur les enregistrements.

Ces deux pistes nécessitent un travail de recherche plus approfondi dans la perspective de diversifier et de complexifier les rendus du synthétiseur. Il s'agit également d'étudier la manière dont ces nouvelles sources sonores peuvent être intégrées dans l'architecture de façon complémentaire.

La possibilité d'intégrer de nouveaux paradigmes souligne le fait que le synthétiseur, dans son architecture actuelle, peut ne pas être suffisant à recréer des arrière-plans totalement contrôlés. L'ajout de samples ou de drones serait un moyen de maîtriser, partiellement, la dépendance du synthétiseur aux enregistrements qu'il utilise. Il s'agit de savoir si l'architecture du synthétiseur peut être améliorée dans son exploitation des enregistrements.

6.2.2 Validations expérimentales

De nouvelles expérimentation seraient nécessaires afin de confirmer les résultats obtenus lors des premiers tests. Ainsi que la littérature le décrit, les tests perceptifs restent le meilleur protocole pour l'évaluation des méthodes de synthèse de textures sonores. Il s'agirait de déterminer plus l'impact qualitatif des différents paramètres de la re-synthèse, pris séparément. La comparaison perceptive avec d'autres méthodes de synthèse, pas uniquement granulaires ou concaténatives, serait aussi intéressante.

D'autre part, le mixage dynamique des différents *Grainers* en parallèle peut donner à entendre des résultats complexes. Il serait donc intéressant de penser un protocole de validation perceptive du rendu du synthétiseur dans son ensemble. Ce protocole pourrait avoir un environnement audio-visuel interactif comme support afin d'évaluer la réactivité du synthétiseur en temps-réel.

6.2.3 Intégrations futures

Le synthétiseur est en cours d'intégration au projet de ville virtuelle interactive Octavia. Il s'agit de mettre en pratique les capacités du synthétiseur à produire des ambiances urbaines dynamiques dans un large environnement virtuel. Ce projet devait permettre l'ajout de nouvelles fonctionnalités au prototype. Dans ce contexte, une première version du synthétiseur est en cours de portage vers le langage C#, ce qui devrait encore faciliter son intégration dans d'autres projets.

Nous avons lié le synthétiseur au moteur de développement Unity3D en utilisant la librairie `libpd4unity`². Ceci nous permet d'intégrer virtuellement le synthétiseur dans des projets développés avec Unity3D. Nous menons actuellement des expérimentations dans ce sens, notamment avec la carte sonore *Paris est un Zoo*³ développée en collaboration avec David Christoffel.

L'architecture ouverte du synthétiseur laisse également entrevoir des usages pour la création d'ambiances qui s'étendent au-delà des arrière-plans sonores urbains. Son intégration en tant qu'outil d'écriture sonore pour des applications plus abstraites peut également être envisagée comme une continuité de ce travail de thèse.

2. <https://github.com/patricksebastien/libpd4unity>

3. <http://www.isothesis.com/sound-design/paris-est-un-zoo.html>

Annexe A

Sonagrammes

Cette annexe présente les sonagrammes référencés dans le chapitre 3. Il s'agit de représentations d'enregistrements d'ambiances que nous avons réalisés à Paris, Liège et San Francisco. Les conditions d'enregistrement sont détaillées dans le chapitre 3.

Les sonagrammes sont réalisés avec le logiciel Acousmographe¹. Afin de les rendre plus lisibles, les sonagrammes sont ici représentés sur la hauteur de la page. L'axe des abscisses représente le temps (la durée de l'enregistrement) et l'axe des ordonnées la distribution des fréquences (le plus souvent de 20 à 20000 hertz).

Le tracé noir représente l'intensité du signal pour chaque fréquence à un moment donné. Plus le tracé est sombre et plus la fréquence associée est forte dans l'enregistrement.

1. <http://www.inagrm.com/accueil/outils/acousmographe>

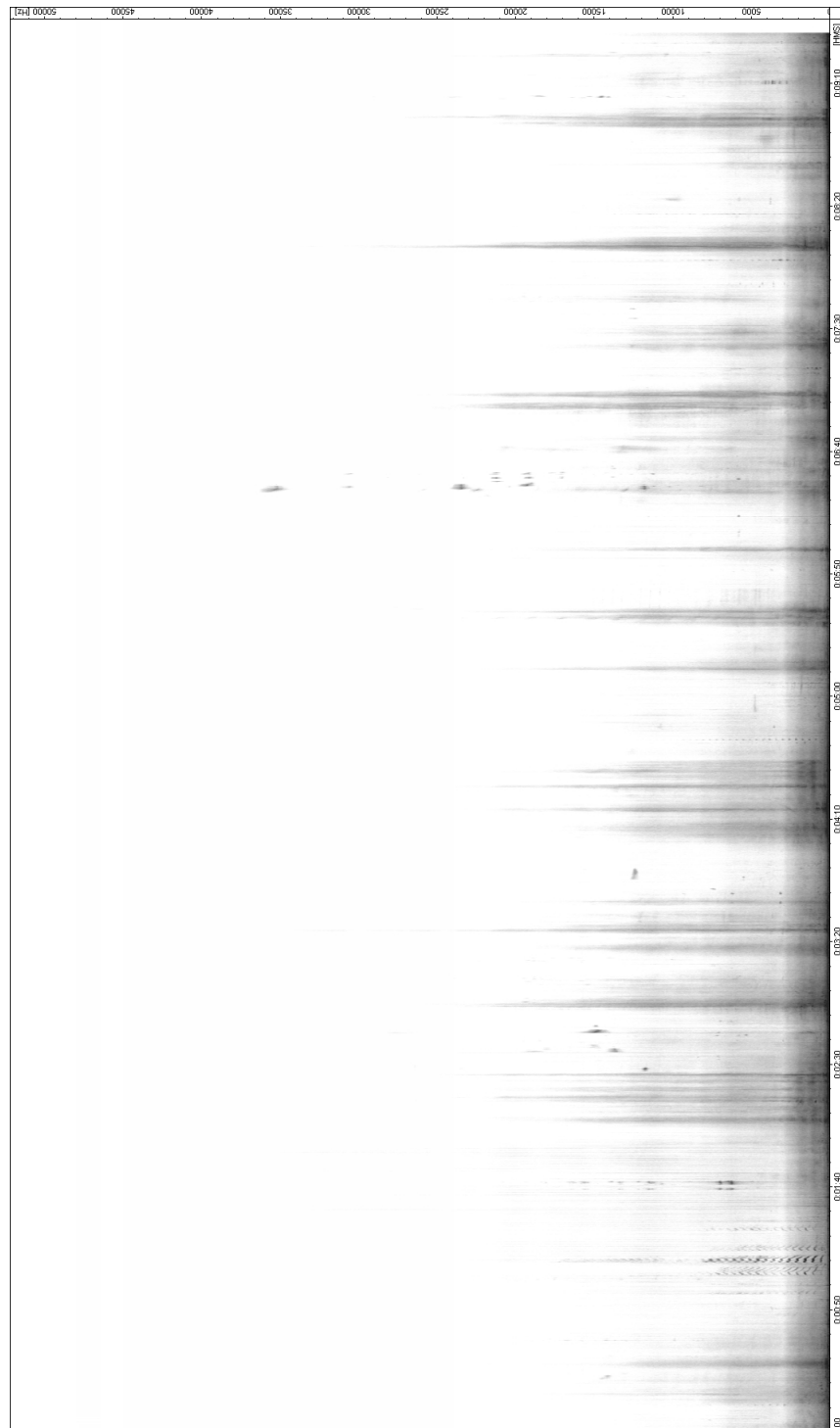


FIGURE A.1 – Sonagramme, promenade sonore de la station Quatre Septembre à Notre Dame de Lorette, Paris

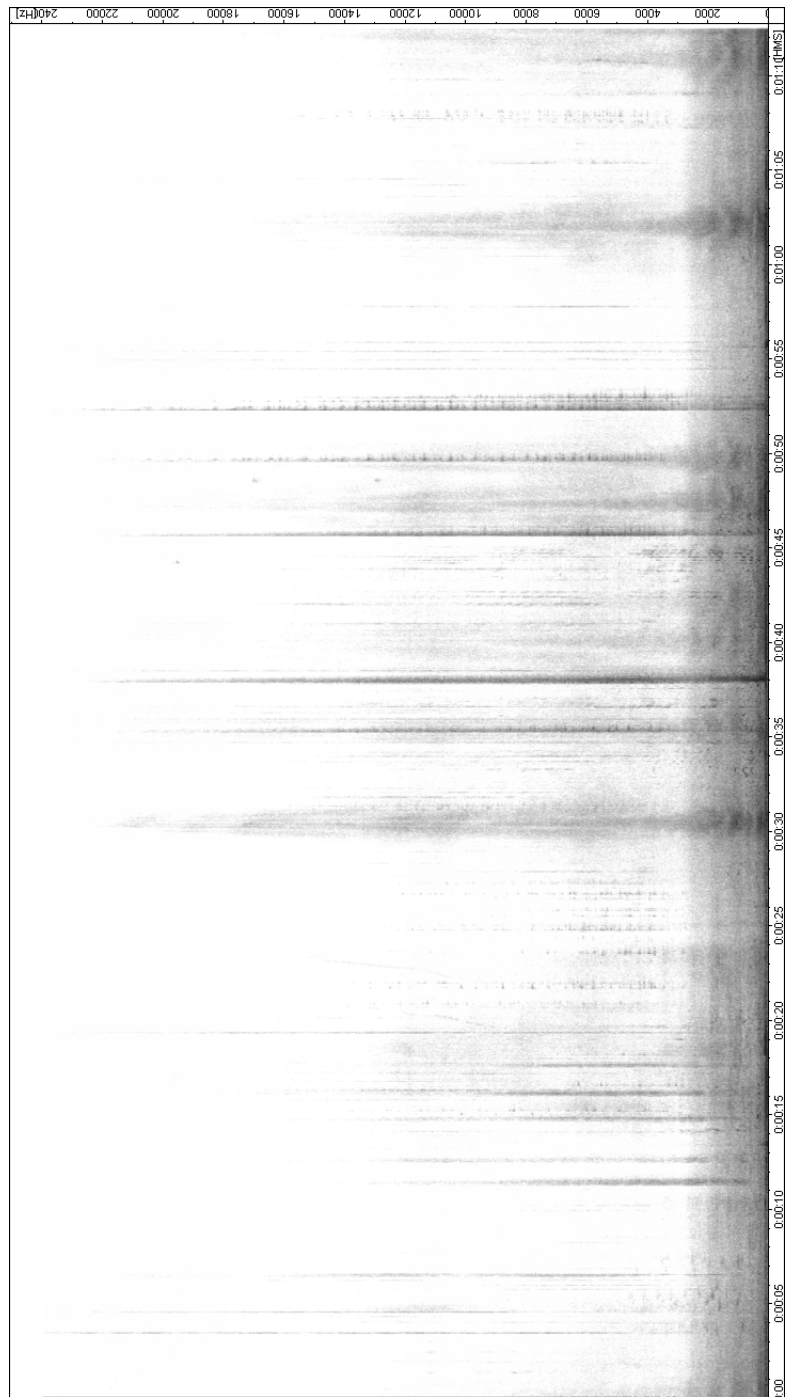


FIGURE A.2 – Sonagramme, ambiance sonore de la ville de Liège 1

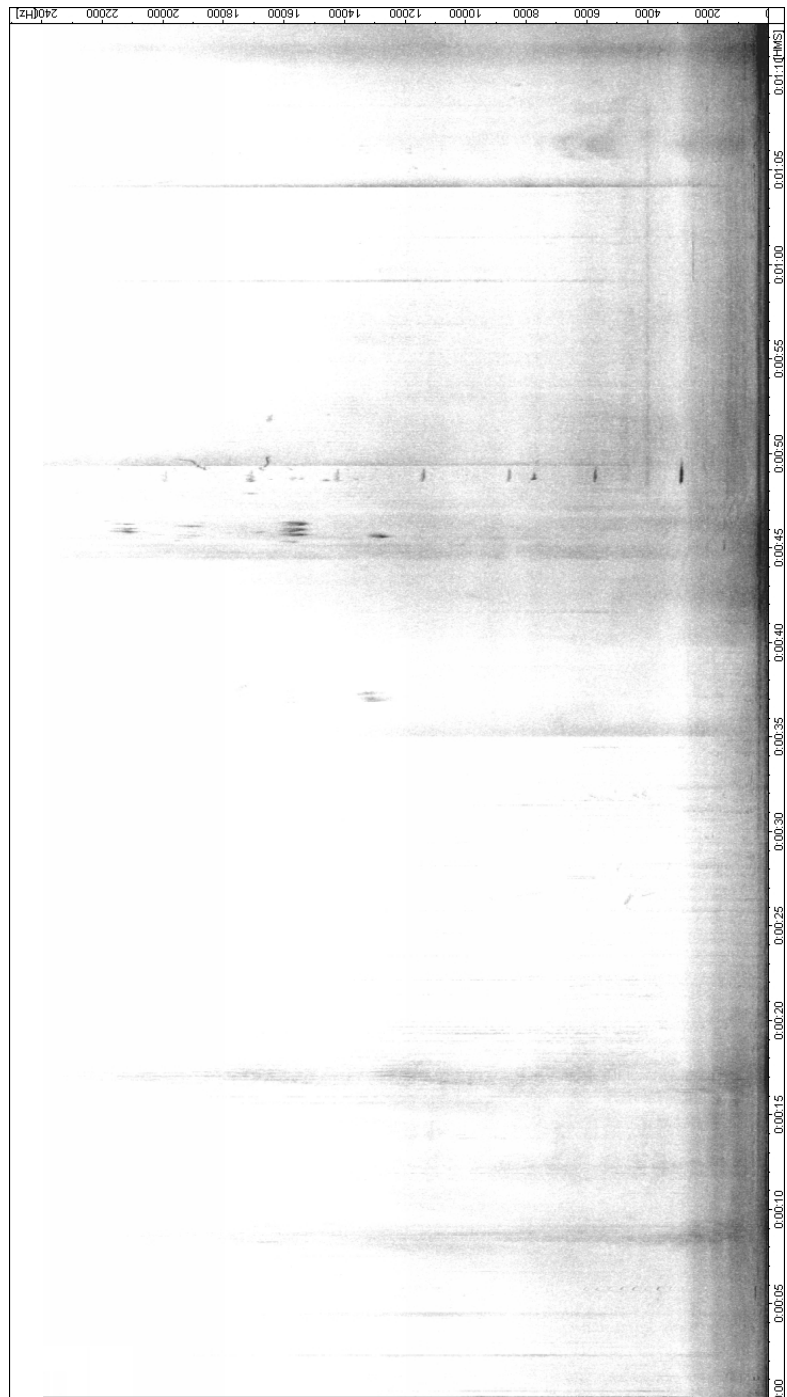


FIGURE A.3 – Sonagramme, ambiance sonore de la ville de Liège 2



FIGURE A.4 – Sonagramme, ambiance sonore de la ville de Liège 3

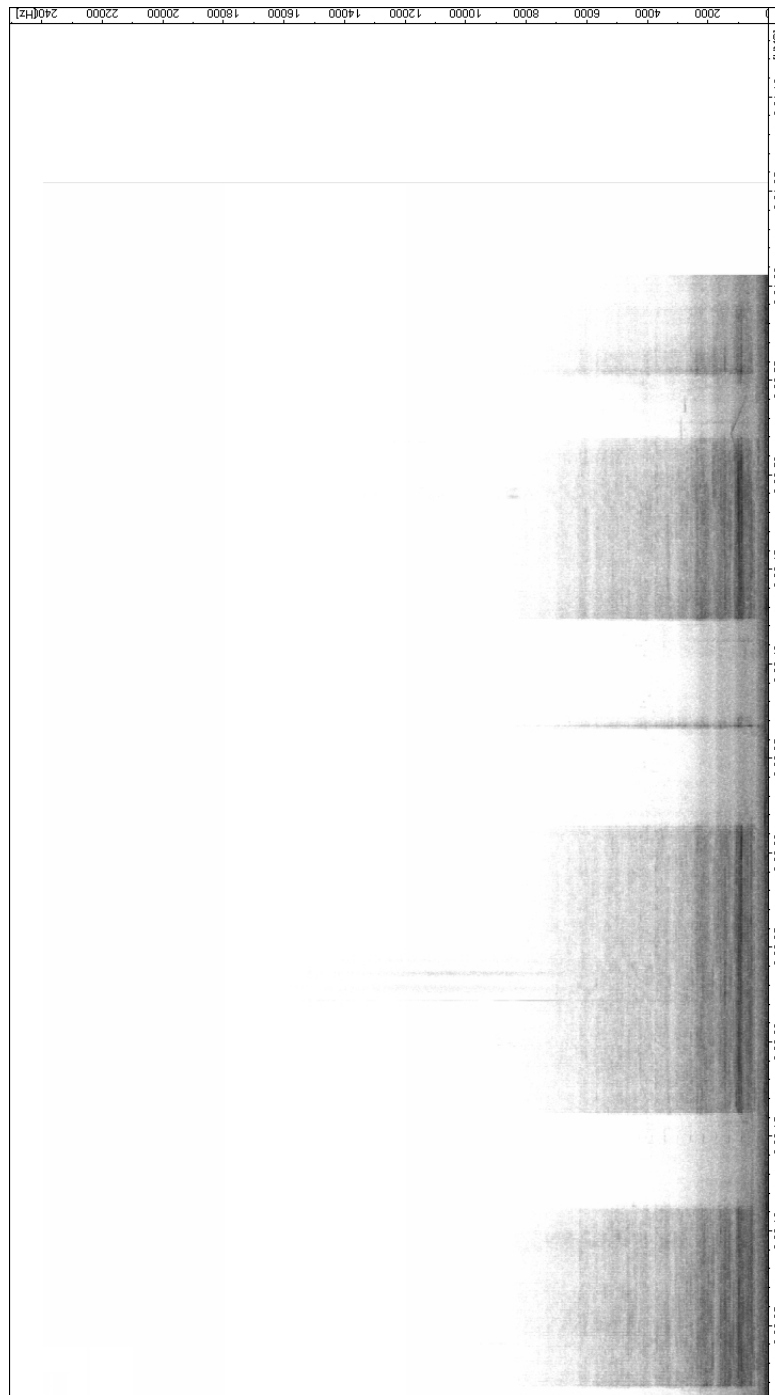


FIGURE A.5 – Sonagramme, ambiance sonore de la ville de Liège 4

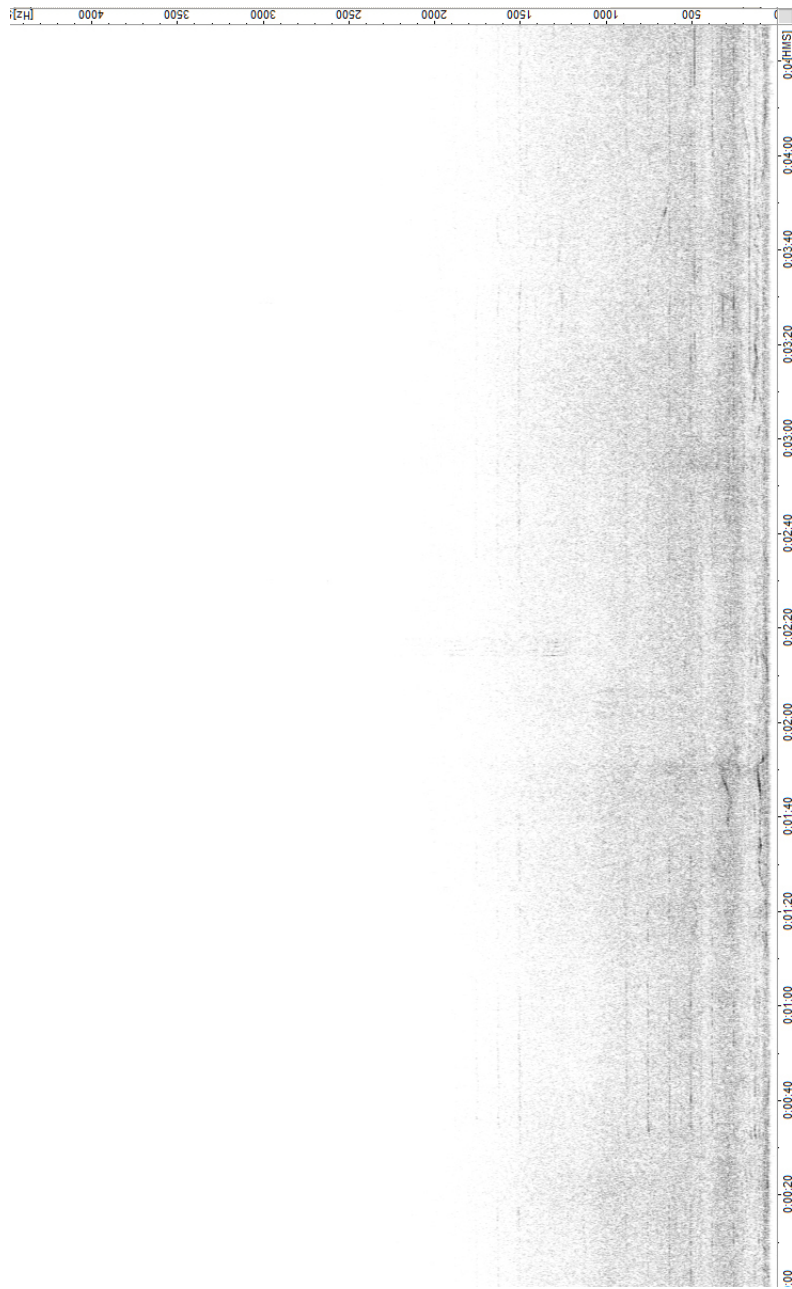


FIGURE A.6 – Sonagramme, ambiance du site BNF captée depuis le parvis de la bibliothèque, Paris

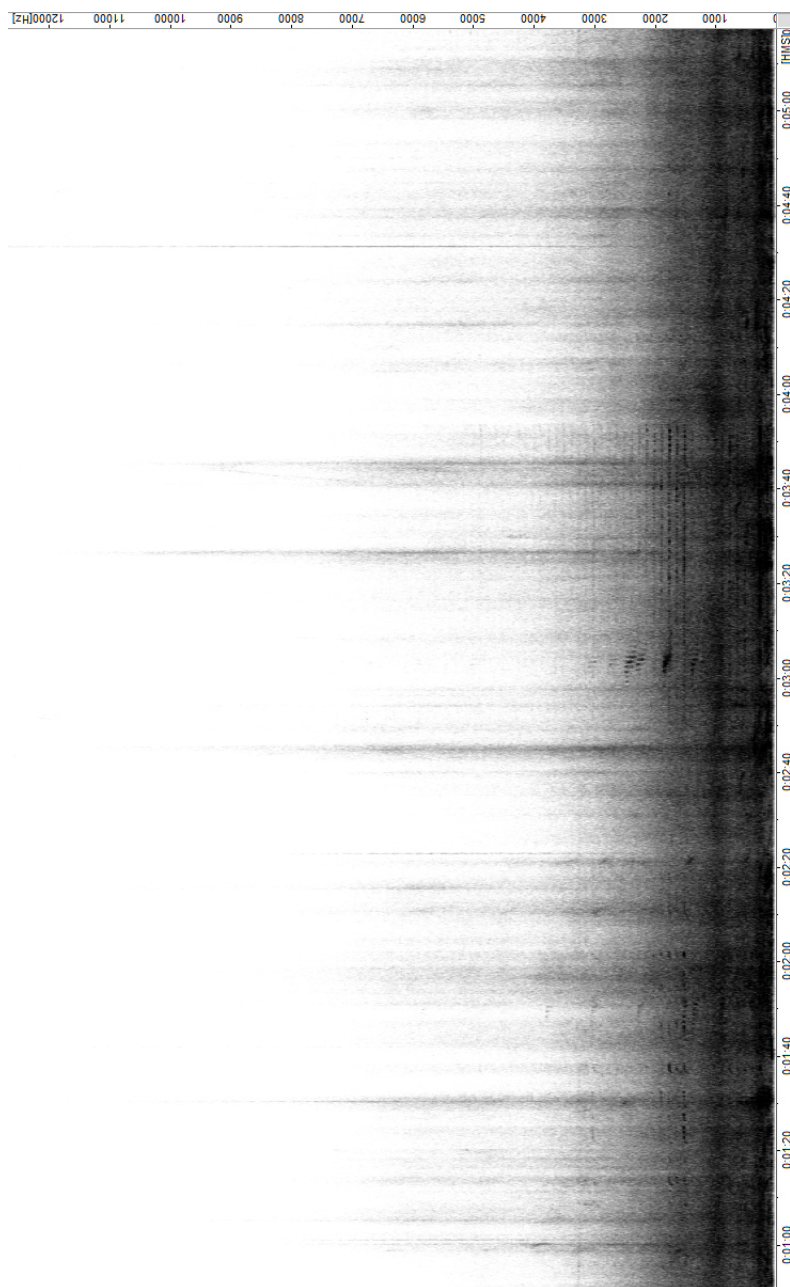


FIGURE A.7 – Sonagramme, ambiance du site BNF captée à proximité du périphérique, Paris

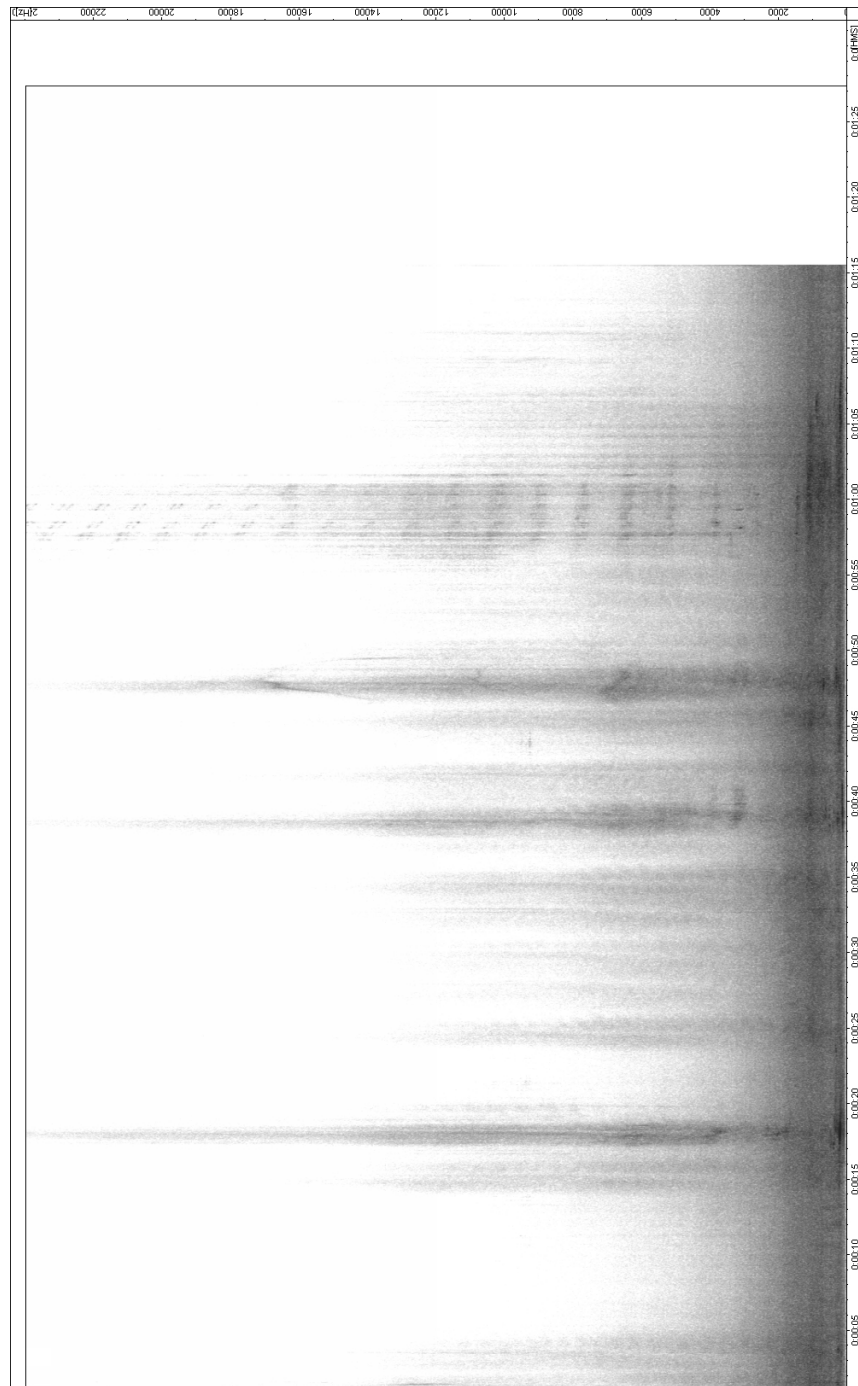


FIGURE A.8 – Sonagramme, San Francisco, ambiance captée au niveau de la circulation en centre ville

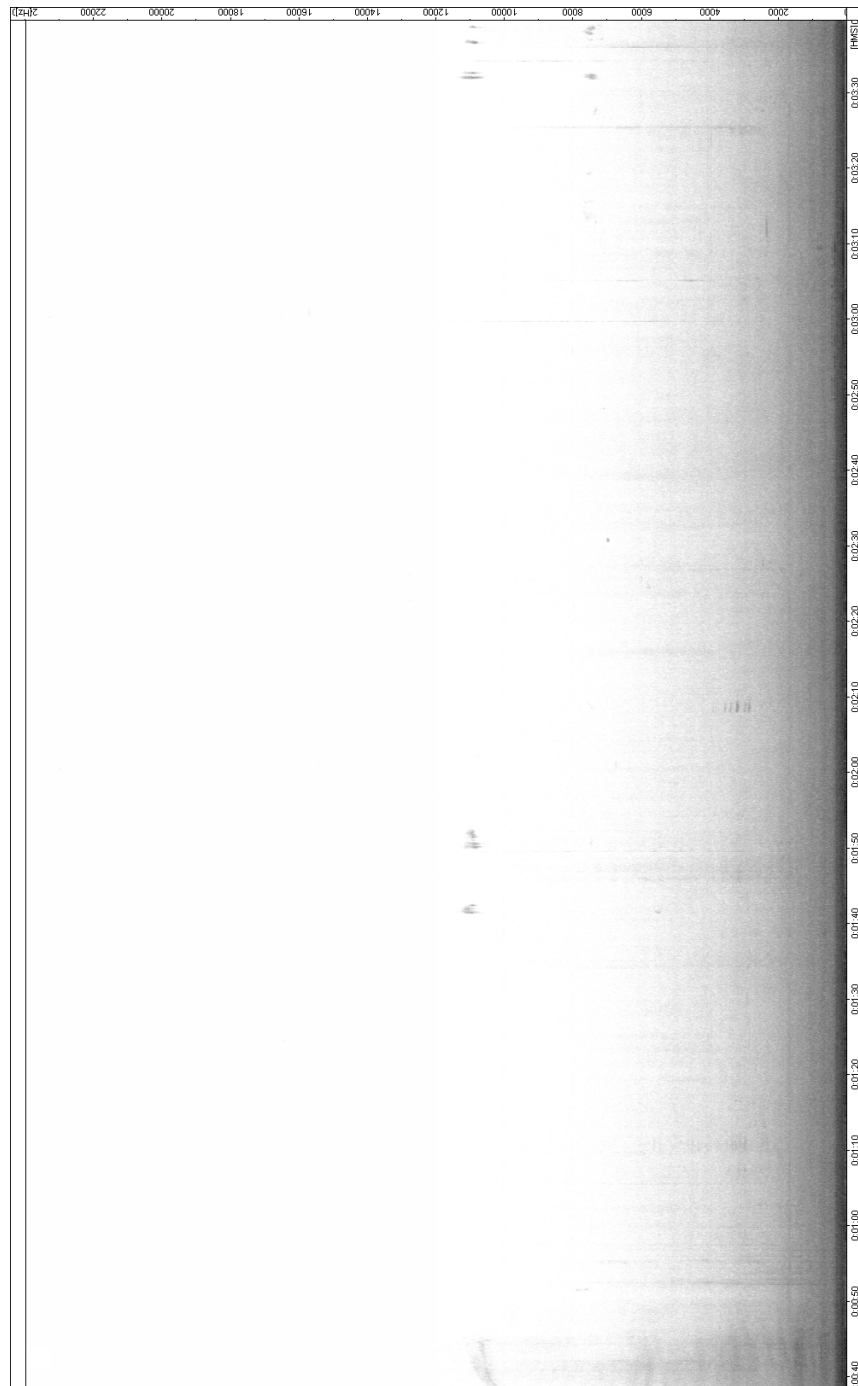


FIGURE A.9 – Sonagramme, San Francisco, ambiance captée depuis le huitième étage d'un parking en centre ville

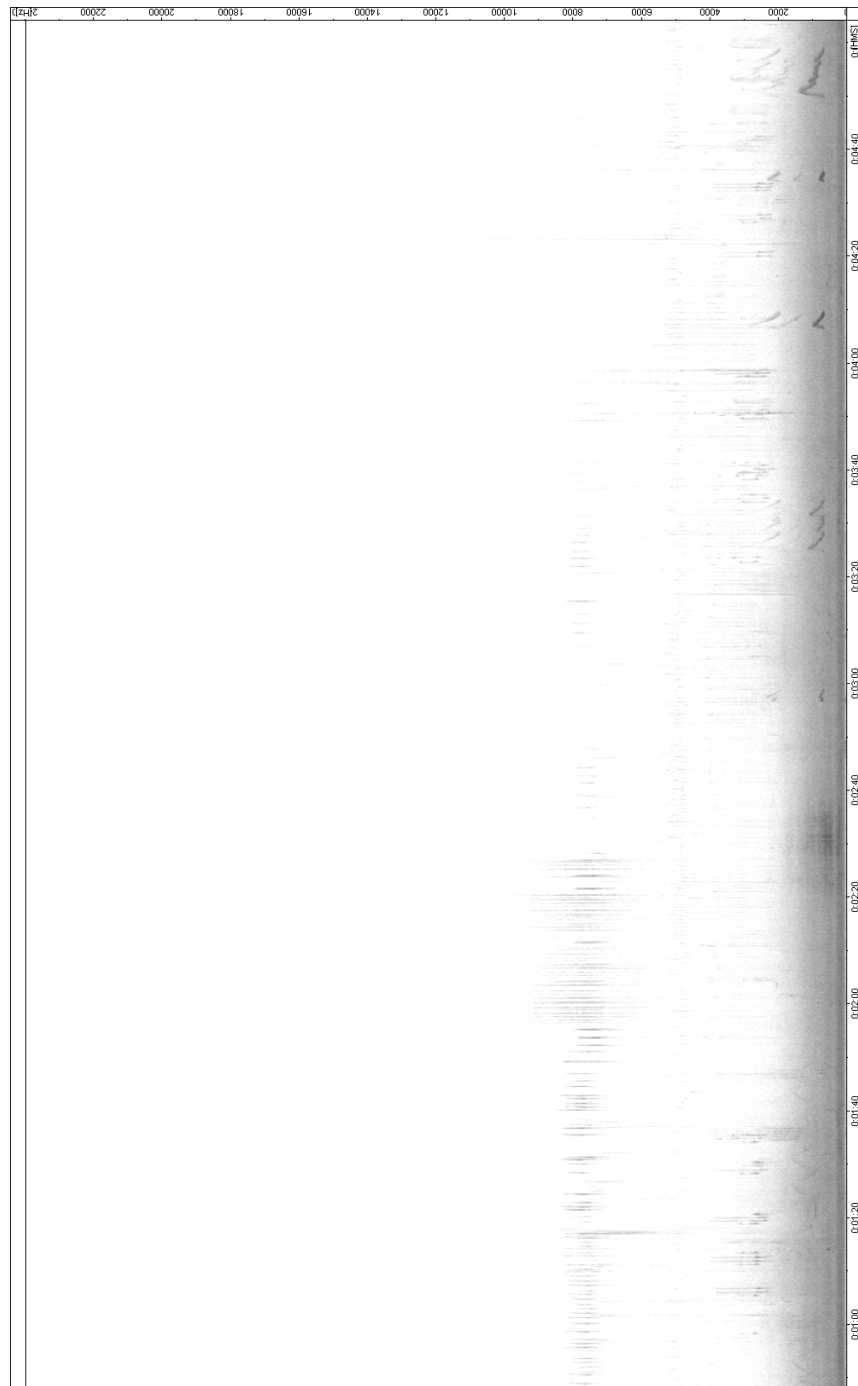


FIGURE A.10 – Sonagramme, San Francisco, ambiance captée depuis les hauteurs du parc Buena Vista, légèrement excentré du centre ville

Annexe B

Exemples de l'utilisation du synthétiseur

Cette annexe présente des exemples de création de textures sonores ambiantes avec le synthétiseur. Il n'existe pas une manière prédéfinie de produire des arrière-plans et des ambiances avec le synthétiseur. Nous souhaitons donc mettre en avant différentes stratégies de sonification selon l'usage fait du synthétiseur. Les interfaces doivent permettre au designer sonore de mettre en place des combinaisons sonores selon des schémas variés.

Dans cette annexe, nous détaillons les modalités d'utilisation du synthétiseur et quelques exemples pour la création d'arrière-plan sonores temps réel complexes et dynamiques. Nous présentons enfin un module que nous avons ajouté au synthétiseur et permettant l'édition dynamique de ces paramètres à l'aide courbes d'automation proches de celles couramment utilisées dans des solutions audio interactives comme Fmod et Wwise.

B.0.4 Ajout de nouveaux fichiers audio

Nous avons souhaité que le synthétiseur soit adaptable à la stratégie de sonification choisie par le designer sonore. Le choix des fichiers audio est le point de départ pour la création de nouvelles textures avec les oscillateurs granulaires. La composition des tables d'ondes granulaires, c'est-à-dire le choix et la distribution des fichiers utilisés par les *Grainers*, peut être thématique ou dynamique, par exemple.

Chacun des quatre *Grainers* d'un oscillateur peut lire un fichier. La principale contrainte à l'utilisation des *Grainers* concerne la taille des fichiers assignés : le nombre de grains disponibles est lié à la fois à la *taille des grains* et à la *taille du fichier* segmenté en grains. Les grains étant lus aléatoirement, le designer sonore devrait prévoir une taille d'échantillon garantissant la création d'au moins une dizaine de grains afin d'éviter d'éventuelles sensations de répétitions.

Les fichiers étant référencés, dans Pure Data, à l'aide de boîtes *messages*, il est possible de charger dynamiquement de nouveaux fichiers dans un lecteur. De plus, nous avons intégré un *sous-patch* permettant de générer un nom complet de fichier ainsi que sa localisation. Cela facilite l'accès à un nouveau contenu dès lors que celui-ci est chargé dans le répertoire dédié et son nom transmis au *sous-patch*. Aucune restriction n'est posée quant au nombre de fichiers audio qu'il est ainsi possible de lier à un lecteur granulaire. Lorsqu'un nouveau son est assigné, celui-ci remplace le précédent dans les tableaux (*arrays*) et donc dans la mémoire vive.

B.0.5 Composition d'une ambiance

L'utilisation du synthétiseur repose sur l'utilisation d'enregistrements d'ambiances utilisés dans un contexte de synthèse granulaire temps réel dynamique.

Nous avons argumenté dans la section 3.4.2 que le choix de la méthodologie des enregistrements n'est pas neutre dans la perception qu'ils donnent à entendre de la ville. Ce choix ressort également au moment de l'intégration des enregistrements dans le synthétiseur. Ainsi, par exemple, l'utilisation d'enregistrements réalisés de différents points d'écoute d'un même lieu peut permettre le mixage dynamique entre ces différents points. Le mixage dynamique des ambiances constitue une difficulté dans l'élaboration d'un environnement virtuel. Le synthétiseur propose certaines facilités pour la composition de différentes ambiances et leur évolution interactive.

Dans les paragraphes suivants, nous proposons différentes stratégies d'utilisation du synthétiseur mettant en lien les méthodologies d'enregistrement des ambiances urbaines avec les fonctionnalités du synthétiseur.

B.0.5.1 Recréer des plans sonores

Dans sa version que nous avons le plus utilisée, le synthétiseur peut utiliser un total de douze fichiers sonores pour composer une ou plusieurs ambiances (quatre fichiers par oscillateur dans un ensemble de trois oscillateurs). Cela laisse au designer sonore une marge stratégique dans la construction des ambiances, selon qu'il souhaite concevoir beaucoup d'ambiances différentes avec peu de fichiers ou, à l'opposé, peu d'ambiances différentes mais avec une granularité plus fine. En tant qu'objets sonores, les ambiances produites avec le synthétiseur sont donc de complexité variable selon le nombre de *Grainers* qui les constituent.

Dans le cas le plus simple, un fichier sonore constitue une ambiance complète, la re-synthèse granulaire met en lecture (et mélange) des fragments sélectionnés aléatoirement depuis cette ambiance pour en faire une texture continue. Douze ambiances différentes pourraient être constituées de cette façon. Le mixage des ambiances a lieu à deux niveaux : à l'intérieur des oscillateur, entre les *Grainers* (voir 4.4) et au niveau du mixeur du synthétiseur, entre les oscillateurs.

Dans le cas le plus complexe, une seule ambiance peut être composée de douze fichiers sonores différents, joués simultanément et mixés ensemble. Les douze fichiers sonores étant répartis entre les trois oscillateurs, l'ambiance complexe peut être scindée en trois sous-parties (par exemple, trafic, foules et sons naturels) et chaque sous-partie est constituée d'un mixage dynamique de quatre fichiers re-synthétisés.

B.0.5.2 Mixage dynamique de différentes ambiances

Des niveaux de complexité intermédiaires permettent d'obtenir des résultats denses et dynamiques. Une ambiance particulière peut être attribuée à chaque oscillateur. Les nuances ou les composantes de chaque ambiance peuvent alors être contrôlées par le mixage dynamique des *Grainers* de chaque oscillateur. Le mixage des trois oscillateurs entre eux permet alors de créer des transitions entre différentes ambiances qui peuvent correspondre à des états de la scène virtuelle ou à des zones géographiques dans l'environnement. Une autre stratégie peut consister à attribuer deux ambiances à deux oscillateurs, respectivement, et utiliser le troisième oscillateur pour la diffusion de sources sonores présentes en complément des deux ambiances. Nous souhaitons que l'ar-

chitecture du synthétiseur soit suffisamment souple pour couvrir une grande diversité d'applications et de stratégies.

Le mixage des *Grainers* d'un oscillateur (voir 4.4.2) peut être mis à profit pour la création de différentes sensations de rendu, notamment des effets de *zoom* ou de progression de densité. Nous avons par exemple effectué des enregistrements à proximité du périphérique parisien, à proximité des voies de circulation, puis en nous éloignant de quelques mètres à chaque nouvel enregistrement. Nous avons ensuite sélectionné quatre prises parmi les enregistrements, les avons éditées et assignées chacune d'entre elles à un *Grainer* d'un oscillateur. Le *Blend Factor* (voir la section 4.4.2) est ajusté de manière à ce que la lecture de plusieurs *Grainers* soit audible simultanément. Le rendu de l'oscillateur pourrait ainsi donner une sensation de rapprochement ou d'éloignement dynamique des voies de circulation, la superposition des différents *Grainers* renforçant le caractère texturé du rendu. Il serait possible d'utiliser le synthétiseur pour agencer des progressions dans des plans sonores plus ou moins proches selon la nature des enregistrements utilisés. De plus, les sensations d'éloignement ou de distance peuvent être renforcées par l'utilisation conjointe des filtre passe-haut et passe-bas.

B.0.6 Courbes d'automation

Nous avons évoqué à plusieurs reprises la nécessité de lier des paramètres de l'environnement virtuel à des paramètres du synthétiseur afin de produire un rendu sonore dynamique. Dans les paragraphes suivants, nous exposons une méthodologie pour utiliser, dans Pure Data, des courbes d'automation afin de transformer la valeur des paramètres entrants.

La normalisation des paramètres choisis est parfois indispensable. Par exemple, le designer sonore souhaite utiliser le nombre de véhicules présents dans une scène virtuelle comme variable entrante d'un paramètre du synthétiseur. Le nombre de véhicules peut varier entre quelques unités et quelques milliers, normaliser ce nombre sur une échelle allant de 0 à 1 rend la variable plus simplement utilisable pour le synthétiseur et permet de définir des valeurs limites. La normalisation peut s'effectuer en temps réel dans Pure Data, sur le flux de données arrivant au paramètre.

Dans la continuité de l'exemple que nous venons de donner, le designer sonore peut souhaiter que l'influence des données entrantes ne soit pas une modification linéaire du paramètre associé ou bien, pour aller plus loin, que plusieurs paramètres liés à une même variable entrante ne soient pas influencés de la même façon. Ces requêtes sont courantes et des moteurs audio comme Fmod proposent de lier un paramètre temps réel à une courbe tracée sur un axe en deux dimension : la variable entrante est normalisée sur l'axe des abscisses et le paramètre associé est modifié en fonction de la valeur en ordonnée correspondante. Le designer sonore choisit la forme de la courbe (linéaire, exponentielle ou bien encore librement tracée) afin de définir de quelle manière la variable entrante influence le paramètre qui lui est associé.

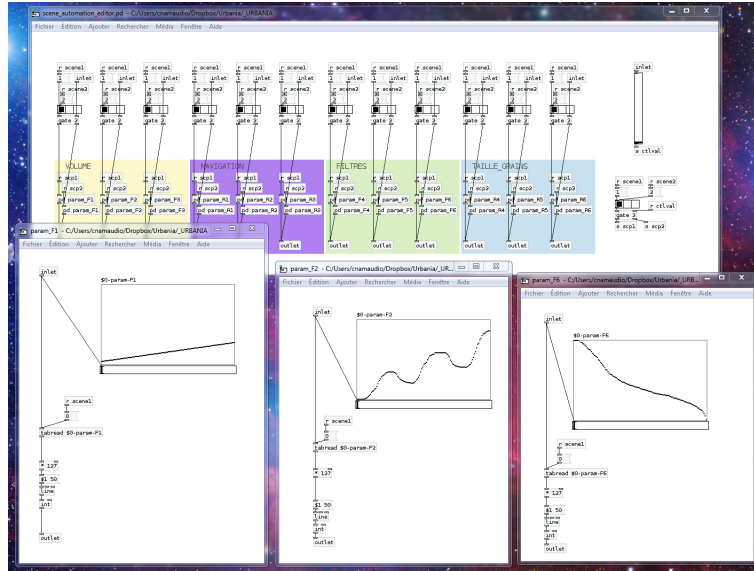


FIGURE B.1 – Exemples de courbes d'automation du synthétiseur d'ambiances urbaines

Nous avons implémenté une solution inspirée de celle de Fmod au synthétiseur réalisé dans Pure Data. La Fig.B.1 montre trois exemples de courbes éditables dans Pure Data. La valeur entrante est considérée en abscisse, la valeur correspondante en ordonnée est ensuite transmise au paramètre de synthèse associé. L'édition de ces courbes permet de faire des correspondances sur-mesure entre une variable entrante et le paramètre qui lui est associé. Ces courbes peuvent ensuite être reliées aux différents paramètres du synthétiseur afin de les faire varier

en temps réel selon l'interprétation qu'elles fournissent des variables entrantes.

L'objectif dans l'utilisation de telles courbes est de pouvoir affiner précisément la réactivité des paramètres les uns par rapports aux autres. Ce type de courbes éditables s'avère souvent indispensable pour donner du sens à l'interprétation d'une variable ou pour créer un mixage dynamique efficace et cohérent.

Annexe C

Expérimentation : choix des échantillons et production des fichiers sonores

Dans cette annexe, nous présentons les détails de la sélection des échantillons de départ utilisés pour l'expérimentation rapportée au chapitre 5 (5.2). Une fois les échantillons sélectionnés pour l'expérimentation, ceux-ci sont re-synthétisés avec *timbreID* et avec le *Grainer* décrit en 4.4.2.



FIGURE C.1 – Sonagramme de l'échantillon A de l'expérimentation

Les Fig.C.1, C.2, C.3 et C.4 sont les sonagrammes des échantillons A, B, C et D choisis pour être re-synthétisés. Dans cette annexe, nous dé-



FIGURE C.2 – Sonagramme de l'échantillon B de l'expérimentation

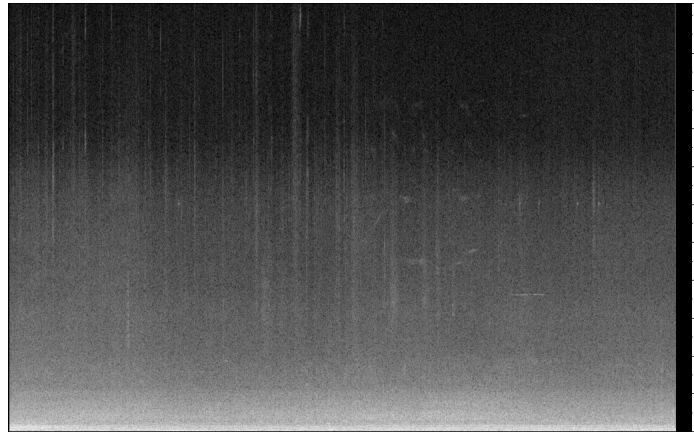


FIGURE C.3 – Sonagramme de l'échantillon C de l'expérimentation

crivons dans un premier temps les fonctionnalités et l'usage fait de *timbreID* pour l'expérimentation. Ensuite, nous détaillons la re-synthèse des échantillons avec le *Grainer*.

C.1 Production des resynthèses avec *timbreID*

TimbreID [Bre10] est un ensemble d'objets compilés pour Pure Data comprenant des descripteurs audio, des outils de synthèse concaténative par corpus et une représentation navigable en deux dimensions



FIGURE C.4 – Sonagramme de l'échantillon D de l'expérimentation

rappelant celle développée pour CataRT [SBVB06]. Les objets dédiés à l'extraction de propriétés de timbre (`magSpec~`, `specBrightness~`, `specCentroid~`, `specFlatness~`, `specFlux~`, `specIrregularity~`, `specKurtosis~`, `specRolloff~`, `specSkewness~`, `specSpread~`, `zeroCrossing~`, `barkSpec~`, `cepstrum~`, `mfcc~`, et `bfcc~`) analysent chaque grain selon une taille de fenêtre définie en samples à l'initialisation du *patch*. Ces analyses peuvent être conduites en temps réel (sur un flux audio) ou sur des fichiers (ou des séries de fichiers) pré-enregistrés. Elles sont ensuite transmises à l'objet `timbreID~` dont la fonction est de stocker les propriétés extraites des analyses et de les distribuer. L'objet `timbreID~` peut procéder à des regroupements parmi les propriétés stockées dans sa mémoire ou parvenir à de nouvelles propriétés relativement à celles déjà stockées. En plus de la classification de timbre selon les propriétés extraites par les descripteurs, cela permet de procéder à des apprentissages automatiques pour classer les grains d'un fichier sonore relativement aux propriétés d'un fichier témoin. D'autres applications sont détaillées dans [Bre10] ainsi que dans les *patches* d'exemple fournis avec *timbreID*¹. Nous utilisons *timbreID* pour décomposer un fichier sonore en grains qui sont ensuite distribués et rejoués en fonction des propriétés extraites de l'analyse de chaque grain.

Afin de produire des fichiers pour l'expérimentation, nous avons modifié le *patch* de démonstration « *timbre-space* ». Ce *patch* propose

1. *TimbreID* est librement téléchargeable depuis le site de son concepteur, William Brent : <http://williambrent.conflations.com/pages/research.html>

un espace en deux dimensions de projection de timbre selon des paramètres d'analyse choisis au préalable. Les propriétés extraites par deux descripteurs sonores prédéfinis permettent, *via* l'objet *timbreID*~, de projeter chaque grain relativement à deux axes. Ce mode de re-synthèse offre un contrôle fin du rendu sonore en parcourant les grains répartis dans l'espace afin d'en déclencher la lecture. Ce *patch* possède une polyphonie de 32 voix, c'est-à-dire que 32 grains peuvent être lus simultanément. Dans cette configuration, l'enchaînement et la superposition des grains sont liés à la navigation manuelle dans l'espace de timbre. La Fig.C.5 montre la fenêtre de configuration de *timbreID* (bas de la Fig.C.5) et la fenêtre de projection de timbre (haut de la Fig.C.5), cette dernière est parcourue à la souris (dont le pointeur est représenté par la zone verte), les grains étant représentés par les sphères de couleurs².

Le *patch* « *timbre-space* » nécessitait quelques ajustements pour produire des fichiers utilisables dans l'expérimentation. Le premier ajustement consistait à rendre le *patch* compatible avec des fichiers stéréo, la version d'origine étant monophonique. Les échantillons que nous avons sélectionnés sont au format stéréo et, comme décrit dans l'état de l'art, les ambiances utilisent majoritairement deux canaux audio, ou davantage si cela est possible. Il paraissait donc important de modifier le *patch* afin de le conformer à ce critère. Par défaut, le *patch* stocke un fichier mono dans un tableau, stocker un fichier stéréo requiert donc de créer un tableau supplémentaire dans lequel stocker le second canal. *TimbreID* analyse le contenu du tableau afin d'organiser les grains (nous décrivons ce processus dans les paragraphes suivants). En raison de l'uniformité des deux canaux sur les échantillons sélectionnés, et du fait qu'il était essentiel de les maintenir liés lors de la lecture, l'analyse est réalisée sur un seul canal.

Le second ajustement concernait la taille des grains utilisés. La taille des grains proposée par *timbreID*, donnée en *samples*, doit être une puissance de deux (1024, 2048, 4096,...). La taille des grains correspond également à la taille de la fenêtre sur laquelle sont effectués les calculs organisant les grains (nous détaillons ces calculs dans le pa-

2. L'attribution des couleurs est ici liée aux résultats de l'analyse des descripteurs de *Bark Frequency Cepstral Coefficients* BFCC1, BFCC2 et BFCC3 assignés respectivement aux valeurs de rouge, vert et bleu. Les BFCC modélisent une distribution de l'énergie spectrale d'un signal de façon perceptivement signifiante. Ces coefficients sont principalement utilisés pour la reconnaissance vocale mais peuvent également trouver des applications dans l'analyse de sons environnementaux [CS02]

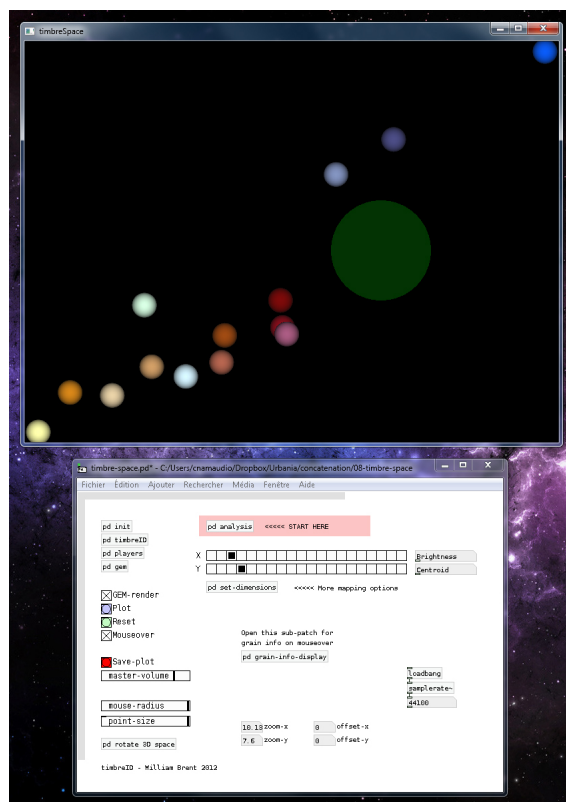


FIGURE C.5 – Pure Data, *timbreID* : fenêtres de configuration (bas) et de projection de timbre (haut) du patch « *timbre-space* ».

ragraphe suivant). Nous avons effectué différents essais de réglage de la taille des grains tout en sachant que nous souhaitons privilégier les grains de grande taille et ce pour une double raison. Les textures sonores de type de celles représentées par les échantillons choisis semblent mieux fonctionner avec de larges grains, le *Grainer* utilise donc des grains assez longs (supérieurs à une seconde) et nous souhaitons, dans un premier temps, utiliser des réglages de granularité proches entre les deux méthodes de resynthèse. La résolution sélectionnée pour *timbreID* est donc de 65536 samples soit 1356 millisecondes. Ce choix donne un compromis entre le nombre de grains répartis dans l'espace de projection de timbre et la durée des grains supérieure à une seconde.

La capacité du patch « *timbre-space* » à organiser les grains dans un espace de projection repose essentiellement sur la connaissance du contenu sonore des grains. Cette connaissance est acquise par le biais

d'une analyse de chaque grain par des descripteurs sonores. Ces descripteurs sont donc une composante essentielle de la synthèse concaténative. Chaque descripteur produit son analyse en accédant au tableau dans lequel l'échantillon sonore est stocké. L'analyse est faite sur des fenêtres correspondant à la taille de grain choisie. Les résultats de l'analyse sont ensuite normalisés entre 0 et 1 avant d'être envoyés à l'objet *timbreID* qui permet d'organiser et de rappeler chaque grain en fonction des valeurs obtenues. La fenêtre de configuration (Fig.C.5, bas) permet de choisir quel descripteur assigner à quel axe de l'espace de projection de timbre (Fig.C.5, haut) afin d'organiser les grains.

Les descripteurs *centroid* et *flatness* semblent être des choix pertinents pour les textures sonores ([Mas11], p. 15) tandis que le descripteur *brightness* s'avère pertinent sur tous les types de sonorités [MML⁺08] et correspond à la sensation de brillance perçue d'un son. Tout comme le descripteur *brightness*, *centroid* fournit des informations sur la distribution de l'énergie spectrale d'un signal, cependant *centroid* recherche la fréquence moyenne du signal. *Flatness* mesure l'uniformité dans la distribution des fréquences sur l'ensemble du spectre ce qui permet de distinguer, aux deux extrêmes, une tonalité pure (une sinusoïde) d'un bruit blanc. Enfin, *amplitude* fait une moyenne de l'amplitude du signal sur la durée de chaque grain.

Si l'efficacité de ces descripteurs sur les textures sonores est donc attestée par la littérature, nous avons tenté différentes configurations des axes afin de parvenir au résultat le plus uniforme dans leur distribution sur l'espace de projection. Nous avons conservé le descripteur *centroid* sur l'axe des ordonnées et utilisé successivement trois descripteurs différents sur l'axe des abscisses : *amplitude*, *brightness* et *flatness*. Ces trois distributions proposent des configurations de grains homogènes, avec les échantillons utilisés, tout en étant suffisamment distinctes les unes des autres. Le parcours de l'espace de projection de timbre offre donc des résultats différents avec chaque configuration des axes.

Les fichiers de re-synthèse sont ensuite produits en parcourant l'espace de projection des grains à l'aide de la souris. Le mélange entre les grains et l'uniformité de la texture résultante sont donc relatifs au déplacement de la souris à dans l'espace. La répartition des grains permet de naviguer entre les différentes nuances de chaque fichier. Notre objectif était d'obtenir le rendu le plus naturel possible, relativement à l'échantillon de départ, et d'utiliser les nuances dans la répartition

spatiale afin de fluidifier l'enchaînement des grains en termes de couleurs.

C.2 Production des resynthèses avec *Grainer*

Comme précisé dans l'introduction de cette section, l'expérimentation ne mobilise pas l'intégralité du synthétiseur mais uniquement un objet *Grainer*. Afin de produire des fichiers de re-synthèse, nous avons donc réalisé un *patch* Pure Data n'utilisant qu'un seul *Grainer* (Fig.C.6).

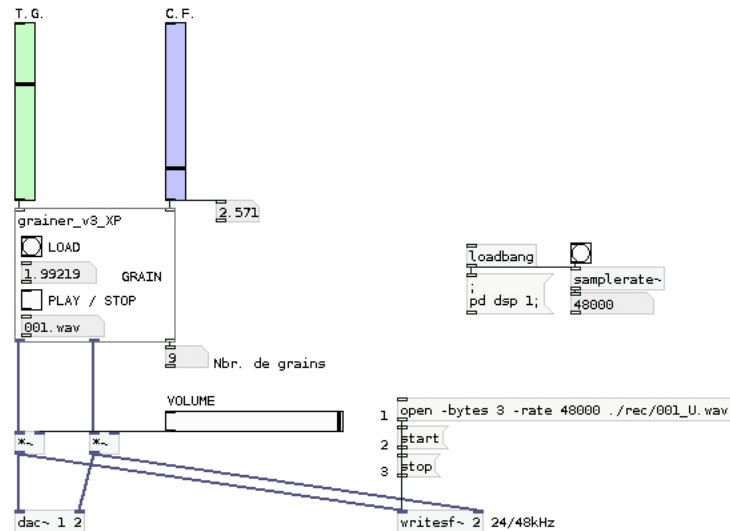


FIGURE C.6 – Pure Data : *patch* utilisé pour générer les fichiers de resynthèse granulaire

Une fois ce *patch* réalisé, la production des fichiers se fait en trois étapes : chargement d'un échantillon dans le *Grainer*, réglages du *Grainer* (taille des grains et ratio de recouvrement entre ceux-ci) et déclenchement de la re-synthèse.

Les réglages choisis sont identiques pour les quatre échantillons, à savoir des grains de 1992 millisecondes avec un temps de recouvrement entre les grains d'environ 775 millisecondes ($1992/2.571$). Tout

comme pour *timbreID*, ces réglages sont choisis après différents essais avec pour objectif d'obtenir le résultat le plus convainquant à l'oreille. La taille des grains n'est donc ici pas la même que celle choisie avec *timbreID*.

Les sorties des *patches timbreID* et *Grainer* sont capturées en temps réel dans Pure Data avec une résolution de 48 kHz et 24 bit. Les fichiers enregistrés sont ensuite édités dans le logiciel *Sound Forge 9.0* afin d'éliminer les temps de silence au début et à la fin de chaque fichier et ainsi d'en réduire la durée à vingt secondes, soit une durée identique à celle des échantillons d'origine.

Questionnaire pour la série A d'échantillons

Pas du tout proche Extrêmement proche

1 2 3 4 5 6 7 8 9 10

Annexe E

Résultats de l'expérimentation

Participant	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3	D4
1	7	9	8	4	9	3	4	3	7	4	7	6	9	7	5	7
2	7	4	5	6	5	5	6	3	7	7	4	5	9	5	4	4
3	4	5	6	6	5	7	6	4	9	5	7	7	5	9	8	6
4	9	6	6	6	7	6	6	6	8	7	7	7	7	8	7	7
5	6	6	6	6	6	7	4	6	8	6	6	7	6	7	7	7
6	9	6	6	6	7	6	7	4	9	6	6	7	6	7	6	10
7	6	7	7	9	6	6	6	7	7	7	6	8	8	7	6	8
8	6	7	5	6	4	5	5	4	7	6	4	4	9	5	6	4
9	6	7	5	6	4	5	5	4	7	6	4	4	9	5	6	4
10	4	4	6	10	4	4	4	4	10	5	4	4	8	4	4	10
11	7	7	6	9	5	6	7	7	9	5	7	6	8	8	7	9
12	6	6	6	7	6	3	6	4	9	6	7	5	8	4	7	7
13	6	6	6	7	6	3	6	4	9	6	7	5	8	4	7	7
14	8	8	8	8	10	8	9	8	8	8	9	7	9	8	7	9
15	6	8	8	7	9	4	5	6	9	4	6	7	9	5	4	8
16	6	7	7	7	10	4	6	4	10	5	7	3	10	7	5	8
17	4	7	7	10	4	4	6	4	10	5	7	4	10	7	5	8
18	4	5	6	9	5	5	3	4	9	5	4	3	9	7	4	6
19	3	3	4	8	2	3	3	2	8	5	3	8	7	3	2	5
20	3	3	4	8	2	3	3	2	8	5	3	8	7	3	2	5
21	8	8	8	9	10	8	7	8	10	8	8	10	8	8	7	8
22	4	6	3	6	5	4	4	5	9	7	4	6	9	5	6	4
23	4	6	3	6	5	4	4	5	9	7	4	6	9	5	6	4
24	3	3	4	8	2	3	3	2	8	5	3	8	7	3	2	5
Moyenne	5,916667	6,041667	6,208333	8,375	5,083333	5,541667	5,041667	5,541667	6,5	5,033333	5,708333	5,583333	6,5	5,791667	5,083333	5,725
Ecartype	1,7424787	1,7315278	1,7440375	1,24455	1,7172454	1,8172882	1,7810384	0,9780193	1,0854019	1,8052861	1,3318085	0,884662	1,5597884	1,6694419	1,863	1,3029109

Annexe F

Résultats des tests t de Student

Comparaison des moyennes A4 et A1 Différence 2,458 t (Valeur observée) 5,624 t (Valeur critique) 2,069 DDL 42 p-value (bilatérale) < 0,0001 alpha 0,05	Comparaison des moyennes B4 et B1 Différence 3,417 t (Valeur observée) 8,469 t (Valeur critique) 2,069 DDL 36 p-value (bilatérale) < 0,0001 alpha 0,05
Comparaison des moyennes A4 et A2 Différence 2,333 t (Valeur observée) 5,361 t (Valeur critique) 2,069 DDL 42 p-value (bilatérale) < 0,0001 alpha 0,05	Comparaison des moyennes B4 et B2 Différence 2,958 t (Valeur observée) 7,023 t (Valeur critique) 2,069 DDL 35 p-value (bilatérale) < 0,0001 alpha 0,05
Comparaison des moyennes A4 et A3 Différence 2,167 t (Valeur observée) 4,954 t (Valeur critique) 2,069 DDL 42 p-value (bilatérale) < 0,0001 alpha 0,05	Comparaison des moyennes B4 et B3 Différence 3,458 t (Valeur observée) 8,338 t (Valeur critique) 2,069 DDL 36 p-value (bilatérale) < 0,0001 alpha 0,05
Comparaison des moyennes C4 et C1 Différence 2,667 t (Valeur observée) 6,863 t (Valeur critique) 2,069 DDL 35 p-value (bilatérale) < 0,0001 alpha 0,05	Comparaison des moyennes D4 et D1 Différence 2,083 t (Valeur observée) 4,881 t (Valeur critique) 2,069 DDL 45 p-value (bilatérale) < 0,0001 alpha 0,05
Comparaison des moyennes C4 et C2 Différence 2,792 t (Valeur observée) 6,803 t (Valeur critique) 2,069 DDL 33 p-value (bilatérale) < 0,0001 alpha 0,05	Comparaison des moyennes D4 et D2 Différence 2,792 t (Valeur observée) 6,298 t (Valeur critique) 2,069 DDL 45 p-value (bilatérale) < 0,0001 alpha 0,05
Comparaison des moyennes C4 et C3 Différence 2,917 t (Valeur observée) 6,725 t (Valeur critique) 2,069 DDL 32 p-value (bilatérale) < 0,0001 alpha 0,05	Comparaison des moyennes D4 et D3 Différence 2,750 t (Valeur observée) 5,597 t (Valeur critique) 2,069 DDL 41 p-value (bilatérale) < 0,0001 alpha 0,05

Bibliographie

- [ABD⁺08] M. Adams, N. Bruce, W. Davies, R. Cain, P. Jennings, A. Carlyle, P. Cusack, K. Hume, and C. Plack. Soundwalking as a methodology for understanding soundscapes. In *Proceedings of the Institute of Acoustics*, volume 30 Pt. 2, 2008.
- [ADB09] M. Adams, B. Davies, and N. Bruce. Soundscapes : an urban planning process map. In *Inter-Noise International Congress on Noise Control Engineering*, 2009.
- [AF02] O. Arikan and D.A. Forsyth. Interactive motion generation from examples. In *SIGGRAPH Š02 : Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 483–490, 2002.
- [And79] U. Andresen. A new way in sound synthesis. In *Audio Engineering Society Convention 62*, 1979.
- [Aro92] B. Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12 :35–50, 1992.
- [ATC04] P. Amphoux, J.-P. Thibaud, and G. Chelkoff. *Ambiances en débat*. 2004.
- [BDD06] D. Botteldooren, B. Decoensel, and T. Demuer. The temporal structure of urban soundscapes. *Journal of Sound and Vibration*, 292(1-2) :105–123, 2006.
- [BDT⁺08] N. Boneel, G. Drettakis, N. Tsingos, I. Viaud-Delmon, and D. James. Fast modal sounds with scalable frequency-domain synthesis. *Association for computing machinery (ACM) Transactions on Graphics (TOG)*, 3 :24, 2008.

- [BJ96] R. Bristow-Johnson. Wavetable synthesis 101, a fundamental perspective. In *Audio Engineering Society Convention 101*, 1996.
- [BMS05] D. Birchfield, N. Mattar, and H. Sundaram. Design of a generative model for soundscape creation. In *International Computer Music Conference (ICMC)*, 2005.
- [Boo04] A. Boothroyd. Room acoustics and speech perception. *Seminars in Hearing*, 25 :155–166, 2004.
- [Bra08] A. Brandon. Next-gen audio square-off : Playstation 3 vs. xbox 360. *Gamasutra*, 2008.
- [Bre94] A.S. Bregman. *Auditory scene analysis : the perceptual organization of sound*. Bradford Books. MIT Press, 1994.
- [Bre10] W. Brent. A timbre analysis and classification toolkit for pure data. In *International Computer Music Conference (ICMC)*, 2010.
- [BRFH⁺11] T. Binder-Reisinger, U. Feldges, H. Hatzmann, M. Offenhuber, and R. Wakolbinger. Notes on artistic research in urban spaces : film, video and sound strategies. *Film and Media Studies*, 4 :89–100, 2011.
- [Bri07a] R. Bridgett. Post-production sound : a new production model for interactive media. *The Soundtrack*, 1 :29–39, 2007.
- [Bri07b] R. Bridgett. Why ambient sounds matter to your game. *Game Developer Magazine*, 2007.
- [BSF00] R. Boulanger, P. Smaragdis, and J. Fitch. Scanned synthesis : an introduction and demonstration of a new synthesis and signal processing technique. In *International Computer Music Conference (ICMC)*, 2000.
- [Che53] E.C. Cherry. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25 :975–979, 1953.
- [Che96] G. Chelkoff. Imaginaire sonore et environnement urbain. *Les cahiers de la recherche architecturale et urbaine*, 38-39 :13, 1996.
- [Chi94] M. Chion. *Audio-vision*. Columbia University Press, 1994.

- [Cho73] J.M. Chowning. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7) :526–534, 1973.
- [Cho89] John M. Chowning. Current directions in computer music research. chapter Frequency modulation synthesis of the singing voice, pages 57–63. MIT Press, Cambridge, MA, USA, 1989.
- [CLLB10] A. Can, L. Leclercq, J. Lelong, and D. Botteldooren. Traffic noise spectrum analysis : dynamic modeling vs. experimental observations. *Applied Acoustics*, 71(8) :764–770, 2010.
- [CNK09] S. Chu, S. Narayanan, and C.-C.J. Kuo. Environmental sound recognition with time-frequency audio features. *IEEE Transactions on Audio, Speech and Language Processing*, 17(6) :1142–1158, 2009.
- [CNTT12] S.-H. Chan, S. Natkin, G. Tiger, and A. Topol. Extensible sound description in collada : a unique file for a rich sound design, 9th advances in computer entertainment technology conference (ace 2012). In *9th Advances in Computer Entertainment Technology Conference (ACE 2012)*, pages 151–166, 2012.
- [Col07] K. Collins. *Essays on sound and vision*, chapter An introduction to the participatory and non-linear aspects of video games audio, pages 263–298. Helsinki University Press, 2007.
- [Col08] K. Collins. *Game sound : an introduction to the history, theory, and practice of video game music and sound design*. The MIT Press, 2008.
- [Col09] K. Collins. An introduction to procedural music in video games. *Contemporary Music Review*, 28(1) :5–15, 2009.
- [Coo97] P. Cook. Physically informed sonic modeling (phism) : synthesis of percussive sounds. *Computer Music Journal*, 21(3) :38–49, 1997.
- [CS02] M. Cowling and R. Sitte. Recognition of environmental sounds using speech recognition techniques. *Advanced Signal Processing for Communications Systems*, 703 :31–46, 2002.

- [Dan01] J. Daniel. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université Pierre et Marie Curie (UPMC), Paris 6, Paris, France, 2001.
- [Del09] B. Delaune. Cartes postales sonores, cinéma pour l'Œoreille, ambient music. esthétiques et procédés picturaux dans les musiques contemporaines, pop-rock et expérimentales. Online, 2009.
- [DGR06] D. Dubois, C. Guastavino, and M. Raimbault. A cognitive approach to urban soundscapes : using verbal data to access everyday life auditory categories. *Acta Acustica united with Acustica*, 92 :865–874, 2006.
- [DMW94] S. Donnadieu, S. McAdams, and S. Winsberg. Caractérisation du timbre des sons complexes. i : Analyse multidimensionnelle. *Journal de Physique IV, Colloque C5, supplément au Journal de Physique III*, 4 :593–596, 1994.
- [DP06] R. Dias Poblete. Manipulation of audio in the wavelet domain processing, a wavelet stream using pd. Master's thesis, Institut für Elektronische Musik (IEM), Graz, Austria, 2006.
- [DPL⁺07] B. Defréville, P. Philippe, C. Lavandier, P. Roy, and Y. François. Objective representation of urban soundscape : application to a parisian neighbourhood. In *Inter-Noise 3rd Southern Europe Workshop on Noise Mapping and Environmental Acoustics*, 2007. 3rd Southern Europe Workshop on Noise Mapping and Environmental Acoustics.
- [DRRP06] B. Defréville, P. Roy, C. Rosin, and F. Pachet. Automatic recognition of urban sound sources. In *Proceedings of the 120th AES Conference, Audio Engineering Society*, pages 1–9. Citeseer, 2006.
- [EL99] A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *The Proceedings of the 7th IEEE International Conference on Computer Vision*, volume 2, 1999.
- [EZM11] D.P.W. Ellis, X. Zeng, and J.H. McDermott. Classifying soundtracks with audio texture features. In *IEEE In-*

- ternational Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, pages 5389–5392, May 2011.
- [Far07a] A. Farnell. An introduction to procedural audio and its application in computer games. In *Audio Mostly Conference*, 2007.
- [Far07b] A. Farnell. Marching onwards : Procedural synthetic footsteps for video games and animation. In *Pd Convention 2007*, 2007.
- [Far07c] A. Farnell. Synthetic game audio with pure data. In *2nd Audio Mostly Conference*. Fraunhofer Institute and Technische Universität Ilmenau, 2007.
- [Far10] A. Farnell. *Designing sound*. The MIT Press, 2010.
- [FG04] J. Friberg and D. Gärdenfors. Audio games : new perspectives on game audio. In *Association for computing machinery (ACM) SIGCHI International Conference on Advances in computer entertainment technology*, pages 148–154, 2004. Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology.
- [Fin09] N. Finney. Autonomous generation of soundscapes using unstructured sound databases. Master’s thesis, Department of Information and Communication Technologies Universitat Pompeu Fabra, Barcelona, Spain, 2009.
- [Fis77] L. Fischer. Enthusiasm : from kino-eye to radio-eye. *Film Quarterly*, 31(2) :25–34, 1977.
- [FJ10] N. Finney and J. Janer. Soundscape generation for virtual environments using community-provided audio databases. In *W3C Workshop : Augmented Reality on the Web*, 2010.
- [Fou10] N. Fournel. Procedural audio for video games : are we there yet ? Online, 2010.
- [Gab47] D. Gabor. Acoustical quanta and the theory of hearing. *Nature*, 4044 :591–594, 1947.
- [Gau08] T. Gaudy. *Etude et développements de jeux vidéo sonores accessibles aux personnes aveugles*. PhD thesis, Conservatoire National des Arts et Métiers (CNAM), Paris, France, 2008.

- [Gav93] W.W. Gaver. What in the world do we hear? an ecological approach to auditory event perception. *Ecological Psychology*, 5 :1–29, 1993.
- [GKW07] B. Gygi, G.R. Kidd, and C.S. Watson. Similarity and categorization of environmental sounds. *Perception & Psychophysics*, 69(6) :839–855, 2007.
- [GLCB07] C. Guastavino, V. Larcher, G. Catusseau, and P. Bousard. Spatial audio quality evaluation : comparing transaural, ambisonics and stereo. In *13th International Conference on Auditory Display (ICAD)*, Montreal, Canada, 2007.
- [Gon08] A. Gonot. *Conception et évaluation d’interfaces de navigation dans les environnements sonores 3D*. PhD thesis, Conservatoire National des Arts et Métiers (CNAM), Paris, France, 2008.
- [Gra95] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2) :50–61, 1995.
- [Gre97] J.M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61 :1270–1277, 1997.
- [Gri08] M. Grimshaw. Sound and immersion in the first-person shooter. *Games Computing and Creative Technologies : Journal Articles (Peer-Reviewed)*, 1 :Paper 3, 2008.
- [Gri10] T. Grill. Re-texturing the sonic environment. In *5th Audio Mostly Conference*, 2010.
- [Gua06] C. Guastavino. The ideal urban soundscape : investigating the sound quality of french cities. *Acta Acustica united with Acustica*, 92 :945–951, 2006.
- [Gua07] C. Guastavino. Categorization of environmental sounds. *Canadian Journal of Experimental Psychology*, 61(1) :54, 2007.
- [Gue70] M.A. Guerzon. The principles of quadraphonic recording. *Studio Sound*, 12 :338–342, 1970.
- [Heu09] K. Heutschi. Calculation of reflections in an urban environment. *Acta Acustica united with Acustica*, 95(4) :644–652, 2009.

- [HP01] R. Hoskinson and D. Pai. Manipulation and resynthesis with natural grains. In *International Computer Music Conference (ICMC)*, 2001.
- [ICD97] F. Iovino, R. Caussé, and R. Dudas. Recent work around modalys and modal synthesis. In *International Computer Music Conference (ICMC)*, 1997.
- [IO05] M. Ismail and D. Oldham. A scale model investigation of sound reflection from building facades. *Applied Acoustics*, 66(2) :123–147, 2005.
- [JBM13] A. Jönsson, R. Breslin, and M. Ma. The ambience table : a serious gaming interface for aiding sound design. In *Serious Games Development and Applications*, pages 151–164, 2013.
- [JLP99] J.-M. Jot, V. Larcher, and J.-M. Pernaux. A comparative study of 3-d audio encoding and rendering techniques. In *Audio Engineering Society Conference : 16th International Conference : Spatial Sound Reproduction*, 1999.
- [Jør07] K. Jørgensen. On transdiegetic sounds in computer games. *Digital Aesthetics and Communication*, 5 :105–117, 2007.
- [Jør11] K. Jørgensen. *Game sound technology and player interaction : concepts and developments*, chapter Time for new terminology ? Diegetic and non-diegetic sounds in computer games revisited, pages 78–97. IGI Publications, 2011.
- [Kan00] J. Kang. Sound propagation in street canyons : comparison between diffusely and geometrically reflecting boundaries. *Journal of acouctic society of America*, 107(3) :1394–1404, 2000.
- [Kan02] J. Kang. Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries. *Journal of Sound and Vibration*, 258(5) :793–813, 2002.
- [KB01] D. Keller and J. Berger. Everyday sounds : synthesis parameters and perceptual correlates. In *8th Brazilian Symposium on Computer Music (SBCM)*, 2001.
- [KHP⁺13] D. Korpi, T. Heittola, T. Partala, A. Eronen, A. Mesáros, and T. Virtanen. On the human ability to discri-

- minate audio ambiances from similar locations of an urban environment. *Personal and Ubiquitous Computing*, 17(4) :761–769, 2013.
- [KJ09] K. Kaliski and D. Joyce. Characterizing soundscapes using spectrograms from long-term third octave band data. In *Inter-Noise 38th International Congress on Noise Control Engineering*, 2009.
- [KM88] R. Kronland-Martinet. The wavelet transform for analysis, synthesis, and processing of speech and music sounds. *Computer Music Journal*, 12(4) :11–20, 1988.
- [KM10] E. Kekou and M. Marangoni. A new sense of city through hearing and sound. In *10th Amber Conference, AMBER art and technology festival*, Istanbul, Turkey, 2010.
- [KMW94] J. Krimphoff, S. McAdams, and S. Winsberg. Caractérisation du timbre des sons complexes. ii : Analyses acoustiques et quantification psychologique. *Journal de Physique IV, Colloque C5, supplément au Journal de Physique III*, 4 :625–628, 1994.
- [Kra87] B.L. Krause. Bioacoustics, habitat ambience in ecological balance. *Whole Earth Review*, 57 :14–18, 1987.
- [Kra02] B.L. Krause. *Wild soundscapes : discovering the voice of the natural world*. Wild Sanctuary Books, 2002.
- [Kro08] D. Kromand. Sound and the diegesis in survival-horror games. In *Audio Mostly Conference*, pages 16–19, 2008.
- [KT98] D. Keller and B. Truax. Ecologically based granular synthesis. In *International Computer Music Conference (ICMC)*, 1998.
- [KVB01] A. Kuijpers and G. Van Blokand. Tyre-road noise models in the last two decades. In *Inter-Noise International Congress on Noise Control Engineering*, 2001.
- [Lab11] B. Labelle. *Acoustic territories*. Continuum, 2011.
- [Laf07] G.-H. Laffont. Urbaphobie et cinéma : le cas jacques tati. In *Ville mal aimée, ville à aimer*, Cerisy-La- Salle, 2007. Colloque de Cerisy-La-Salle.
- [LGMV07] S. Le Groux, J. Manzolli, and P.F.M.J. Verschure. Interactive sonification of the spatial behavior of human and

- synthetic characters in a mixed-reality environment. In *10th Annual International Workshop on Presence*, Barcelona, Spain, 2007.
- [LN10] S. Luzzi and R. Natale. Soundscapes in the participatory design of florentine quiet areas. In *17th International Congress on Sound and Vibration (ICSV17)*, 2010.
- [LNT12] G. Levieux, S. Natkin, and A. Topol. Out of context augmented navfields : designing crowd choreographies. In *Entertainment Computing - ICEC 2012*, pages 326–332, 2012.
- [LP13] C. Le Prado. The listening walker : interactive sound walk in a virtual city. In *Entertainment Computing - ICEC 2013*, pages 209–210, 2013.
- [LWZ04] L. Lu, L. Wenyin, and H.-J. Zhang. Audio textures : theory and applications. *IEEE Transactions on Speech and Audio Processing*, 12(2) :156–167, 2004.
- [Lyo74] R.H. Lyon. Role of multiple reflections and reverberation in urban noise propagation. *Journal of acoustic society of America*, 55(3) :493–503, 1974.
- [Mac09] P. Mac. Dead space. *Audio Media*, 224 :28–29, 2009.
- [Mal11] G. Malgouyard. Analyse et synthèse de sons environnementaux. Master’s thesis, Université Pierre et Marie Curie (UPMC) Acoustique, traitement du signal, informatique, appliqués à la musique (ATIAM) filère Systèmes et Applications Répartis (SAR), 2011. Rapport de stage.
- [Mar09] N. Mariette. Ambigrainer - a high order ambisonic granulator in pd. In *Ambisonics Symposium*, 2009.
- [Mas11] A. Masurelle. Towards a gestural control of environmental sound texture synthesis. Master’s thesis, Université Pierre et Marie Curie (UPMC) Acoustique, traitement du signal, informatique, appliqués à la musique (ATIAM), 2011. Master’s Thesis.
- [MBT⁺07] T. Moeck, N. Boneel, N. Tsingos, G. Drettakis, I. Viaud-Delmon, and D. Alloza. Progressive perceptual audio rendering of complex scenes. In *Association for computing machinery (ACM) Symposium on Interactive 3D*

- Graphics and Games (SIGGRAPH)*. ACM SIGGRAPH, 2007.
- [MC05] N.E. Miner and T.P. Caudell. Using wavelets to synthesize stochastic-based sounds for immersive environments. *Association for computing machinery (ACM) Transactions on Applied Perception*, 2 :521–528, 2005.
- [MC09] A. Misra and P.R. Cook. Toward synthesized environments : a survey of analysis and synthesis methods for sound designers and composers. In *International Computer Music Conference (ICMC)*, 2009.
- [McC99] A. McCartney. *Sounding places : situated conversations through the soundscape compositions of Hildegard Westerkamp*. PhD thesis, Graduate Programme in Music, York University, Toronto, Ontario, Canada, 1999.
- [MCD98] V. Maffiolo, M. Castellengo, and D. Dubois. Qualité sonore de l’environnement urbain : sémantique et intensité. *Acoustique et Techniques*, 16 :14–21, 1998.
- [MCW06] A. Misra, P.R. Cook, and G. Wang. A new paradigm for sound design. In *9th International Conference on Digital Audio Effects (DAFx06)*, 2006.
- [MdC12] F. Morais da Costa. Urban sounds and their hearing through cinema. *Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação*, 15(2) :1–14, 2012.
- [Men10] D. Menzies. Physically motivated environmental sound synthesis for virtual worlds. Online, 2010.
- [Mer05] B. Merlier. Réflexions à propos de la mise en espace de la musique électroacoustique dans les logiciels audionumériques. Online, 2005.
- [MH74] J. Malchaire and S. Horstman. Urban noise model. *Journal of acoustic society of America*, 56(6) :1811–1814, 1974.
- [MM95] D.G. Malham and A. Myatt. 3-d sound spatialization using ambisonic techniques. *Computer Music Journal*, 19(4) :58–70, 1995.

- [MMG09] E. Martín Martínez Gutiérrez. Images et imaginaires de la grande ville : variations sur une symphonie urbaine. *Sociétés*, 103(1) :33–46, 2009.
- [MMHS10] A. Minard, N. Misdariis, O. Houix, and P. Susini. Catégorisation de sons environnementaux sur la base de profils morphologiques. In *10ème Congrès Français d'Acoustique*, pages –, Lyon, France, 2010.
- [MML⁺08] A. Minard, N. Misdariis, G. Lemaitre, P. Susini, S. McAdams, and E. Parizet. Environmental sound description : comparison and generalization of 4 timbre studies. In *CHI - Human Factors in Computing Systems Conference*, 2008.
- [MMS⁺10] N. Misdariis, A. Minard, P. Susini, G. Lemaitre, S. McAdams, and E. Parizet. Environmental sound perception : meta-description and modeling based on independent primary studies. Online, May 2010.
- [Mor98] S. Morin. L'écran de la passion, une étude du cinéma de marguerite duras. *Cinémas : Revue d'études cinématographiques*, 8(3) :183–188, 1998.
- [Mor09] S. Morgan. Dynamic game audio ambiance : bringing prototype's new york city to life. *Gamasutra*, 2009.
- [MS10] R. Murray Schafer. *Le paysage sonore : le monde comme musique*. Wildproject, 2010.
- [MS11] J.H. McDermott and E.P. Simoncelli. Sound texture perception via statistics of the auditory periphery : evidence from sound synthesis. *Neuron*, 71 :926–940, 2011.
- [MWD⁺95] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical : common dimensions, specificities, and latent subject classes. *Psychological Research*, 58 :177–192, 1995.
- [Nap08] J. Napolitano. Dead space sound design : in space no one can hear you interens scream. they are dead. (interview). Online, 2008. Interview de Don Veca, Directeur Audio du jeu "Dead Space".
- [NLP07] S. Natkin and C. Le Prado. Listen lisboa : scripting languages for interactive musical installations. In *Sound and Music Computing Conference*, 2007.

- [NLP13] S. Natkin and C. Le Prado. Du son et des jeux vidéo. *INA-SUP : E-dossiers de l'audiovisuel*, Le son dans tous ses états, 2013.
- [Nor84] M.F. Norden. The avant-garde cinema of the 1920s : connections to futurism, precisionism, and suprematism. *Leonardo*, 17(2) :108–112, 1984.
- [O'D02] M. O'Donnell. Producing audio for halo. *Gamasutra*, 2002.
- [OK07] D. O'Regan and A. Kokaram. Multi-resolution sound texture synthesis using the dual-tree complex wavelet transform. In *European Signal Processing Conference (EU-SIPCO)*, 2007.
- [Pau03] L.J. Paul. Audio prototyping with pure data. *Gamasutra*, 2003.
- [Pau11] L.J. Paul. Granulation of sound in video games. In *Audio Engineering Society Conference : 41st International Conference : Audio for Games*, 2 2011.
- [PBA⁺08] J.-D. Polack, J. Beaumont, C Arras, M. Zekri, and B. Robin. Perceptive relevance of soundscape descriptors : a morpho-typological approach. In *Acoustics 08 Paris*, Paris, France, 2008.
- [PFG⁺11] B.C. Pijanowski, A. Farina, S.H. Gage, S.L. Dumyahn, and B.L. Krause. What is soundscape ecology ? an introduction and overview of an emerging new science. *Landscape Ecology*, 26(9) :1213–1232, 2011.
- [PGD12] J. Picaut, G. Guillaume, and G. Dutilleux. Ambiances sonores urbaines et interaction ville-bâtiment. *RéférenceS, Modélisation urbaine : de la représentation au projet* :34–50, 2012.
- [PHS99] J. Picaut, J. Hardy, and L. Simon. Sound propagation in urban areas : a periodic disposition of buildings. *Physical Review E*, 60 :4851–4859, 1999.
- [PS05] I. Pires and N. Schnell. Du grain à la musique : manipulations en temps réel. In *Journées d'informatique musicale JIM05*, 2005.

- [Puc88] M. Puckette. The patcher. In *International Computer Music Conference (ICMC)*, pages 420–429, San Francisco, United-States, 1988.
- [Puc07] M. Puckette. *The theory and technique of electronic music*. World Scientific Press, 2007.
- [PVrD⁺11] B.C. Pijanowski, L.J. Villanueva-rivera, S.L. Dumyahn, A. Farina, B.L. Krause, B.M. Napoletano, S.H. Gage, and N. Pieretti. Soundscape ecology : the science of sound in the landscape. *BioScience*, 61(3) :203–216, 2011.
- [RD92] X. Rodet and P. Depalle. Spectral envelopes and inverse fft synthesis. In *International Computer Music Conference (ICMC)*, 1992.
- [RD05] M. Raimbault and D. Dubois. Urban soundscapes : experiences and knowledge. *Cities*, 22(5) :339–350, 2005.
- [Roa04] C. Roads. *Microsound*. The MIT Press, 2004.
- [Roc99] M. Rocha. *Les techniques granulaires dans la synthèse sonore*. PhD thesis, Université de Paris VIII, 1999.
- [Ros12] E. Ross. Sounds from the city in film noir. Online, 2012.
- [RSM⁺10] N. Raghuvanshi, J. Snyder, R. Mehra, M. Lin, and N. Govindaraju. Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. *Association for computing machinery (ACM) Transactions on Graphics (TOG)*, 29(4), 2010. <http://research.microsoft.com/pubs/131479/paper.pdf> Manque les pages.
- [Rus03] Luigi Russolo. *L’art des bruits, manifeste futuriste 1913*. 2003.
- [RV13] M Rychtáriková and G Vermeir. Soundscape categorization on the basis of objective acoustical parameters. *Applied Acoustics*, 74 :240–247, 2013.
- [San03] U. Sandberg. The multi-coincidence peak around 1000 hz in tyre/road noise spectra. In *Euro-Noise (European Acoustics Association)*, 2003.
- [SAP97] N. Saint-Arnaud and K. Popat. Analysis and synthesis of sound texture. In *Computational Auditory Scene Analysis : Proceedings of the Ijcai-95 Workshop (International*

- Joint Conference on Artificial Intelligence*), pages 293–308, Montreal, Quebec, Canada, 1997. Canadian Society for Computational Studies of Intelligence (CSCSI).
- [SBVB06] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton. Real-time corpus-based concatenative synthesis with cartart. In *9th International Conference on Digital Audio Effects (DAFx06)*, pages 279–282, 2006.
- [SC00] B. Shinn-Cunningham. Learning reverberation : considerations for spatial auditory displays. In *International Conference on Auditory Display (ICAD)*, pages 126–134, 2000.
- [Sch06] D. Schwarz. Concatenative sound synthesis : the early years. *Journal of New Music Research*, 35(1)(1) :3–22, 2006.
- [Sch11] D. Schwarz. State of the art in sound texture synthesis. In *14th International Conference on Digital Audio Effects (DAFx11)*, 2011.
- [Sem06] C. Semidor. Listening to a city with the soundwalk method. *Acta acustica united with acustica*, 92(6) :959–964, 2006.
- [SER06] G. Strobl, G. Eckel, and D. Rocchesso. Sound texture modeling : a survey. In *International Conference on Sound and Music Computing (SMC)*, pages 61–65, 2006.
- [SJB14] J. Salamon, C. Jacoby, and J. P. Bello. A dataset and taxonomy for urban sound research. In *in Proc. 22nd ACM International Conference on Multimedia*, 2014.
- [SMM⁺05] P. Susini, S. McAdams, N. Misdariis, G. Lemaitre, and S. Winsberg. Timbre des sons environnementaux. In *Colloque International de Musicologie*, 2005.
- [SMW99] P. Susini, S. McAdams, and S. Winsberg. A multidimensional technique for sound quality assessment. *Acta Acustica united with Acustica*, 85(5) :650–656, 1999.
- [SMWM98] P. Susini, N. Misdariis, S. Winsberg, and S. McAdams. Caractérisation perceptive de bruits. *Acoustique et Techniques*, 13 :11–15, 1998.

- [SO72] E.A. Shaw and N. Olson. Theory of steady-state urban noise for an ideal homogeneous city. *Journal of the Acoustical Society of America*, 51(6) :1781–1793, 1972.
- [SS10] D. Schwarz and N. Schnell. Descriptor-based sound texture sampling. In *International Conference on Sound and Music Computing (SMC)*, 2010.
- [Sto03] A. Stockburger. The game environment from an auditive perspective. In *Level Up, Digital Games Research Conference*, 2003.
- [Str07] G. Strobl. Parametric sound texture generator. Master’s thesis, Institut für Elektronische Musik und Akustik (IEM) Universität für Musik und darstellende Kunst, Technische Universität, Graz, Austria, 2007.
- [SW98] M.J. Singer and B.G. Witmer. Measuring presence in virtual environments : a presence questionnaire. *Presence*, 7 :225–240, 1998.
- [TFNC01] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom. Modeling acoustics in virtual environments using uniform theory of diffraction. In *Association for computing machinery (ACM) Symposium on Interactive 3D Graphics and Games (SIGGRAPH)*, 2001.
- [TGD04] N. Tsingos, E. Gallo, and G. Drettakis. Perceptual audio rendering of complex virtual environments. *Association for computing machinery (ACM) Symposium on Interactive 3D Graphics and Games (SIGGRAPH) Transactions on Graphics*, 23(3) :249–258, 2004. ACM Transactions on Graphics (SIGGRAPH Conference Proceedings).
- [The12] M. Theiler. Creating ambient audio for la noire and the banner saga. *Gamasutra*, 2012.
- [Thi02] J.-P. Thibaud. From situated perception to urban ambiances. In *International Workshop on Architectural and Urban Ambient Environment*, 2002.
- [Tho10] M. Thomsen. Designing for immersion : recreating physical experience in games. *Gamasutra*, 2010.
- [Tor05] H. Torgue. Immersion et émergence : qualités et significations des formes sonores urbaines. *Espaces et sociétés*, 4(122) :157–166, 2005.

- [Tru78] B. Truax. *The handbook of acoustic ecology*. A.R.C. Publications, 1978.
- [Tru84] B. Truax. *Acoustic communication*. Ablex Publishing Corporation, 1984.
- [Tru90] B. Truax. Composing with real-time granular sounds. *Perspectives of New Music*, 28 :120–134, 1990.
- [Tru92] B. Truax. Composing with time-shifted environmental sounds. *Leonardo Music Journal*, 2 :37–40, 1992.
- [Tru08] B. Truax. Soundscape composition as global music : electroacoustic music as soundscape. *Organised Sound*, 13(2) :103–109, 2008.
- [VAKMP10] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. A 3d immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18 :1550–1561, 2010.
- [VdDKP01] K. Van den Doel, P.G. Kry, and D.K. Pai. Foleyautomatic : physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques*, pages 537–544, 2001. Proceedingd of the 28th annual conference on computer graphics and interactive techniques pp537-544.
- [VdDP03] K. Van den Doel and D.K. Pai. Modal synthesis for vibrating objects. Online, 2003.
- [Ven09] O. Veneri. *Architecture d’un intergiciel pour la création sonore dans les jeux vidéo*. PhD thesis, Conservatoire National des Arts et Métiers (CNAM), Centre d’Etude et de Recherche en Informatique et communications (CEDRIC), Paris, France, 2009.
- [Ver03] V. Verfaillie. *Utilisation d’espaces perceptifs pour la synthèse et la transformation sonore*, chapter Espaces sonores - actes de recherche, pages 39–54. CICM, 2003.
- [Ver10] C. Verron. *Synthèse immersive de sons d’environnement*. PhD thesis, Université de Provence Aix-Marseille I, Aix-Marseille, France, 2010.

- [VMS01] B. Verplank, M. Mathews, and R. Shaw. Scanned synthesis. *Journal of acouctic society of America*, 105 :2400–2404, 2001.
- [Vog99] C. Vogel. *Etude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain*. PhD thesis, Université Pierre et Marie Curie (UPMC), Paris 6, Paris, France, 1999.
- [VRPDP11] L.J. Villanueva-Rivera, B.C Pijanowski, J. Doucette, and B. Pekin. A primer of acoustic analysis for landscape ecologists. *Landscape Ecolology*, 26 :1233–1246, 2011.
- [VS06] F. Venot and C. Sémidor. The "soundwalk" as an operational component for urban design. In *23rd Conference on Passive and Low Energy Architecture (PLEA)*, Geneva, Switzerland, 2006.
- [Wes99] H. Westerkamp. Soundscape composition : linking inner and outer worlds. In *Soundscape Before 2000*, Amsterdam, Nederland, 1999.
- [Wie64] N Wiener. Spatio-temporal continuity, quantum theory and music. In *The Concepts of Space and Time*, pages 539–546. D. Reidel Publishing Company, 1964.
- [WMG65] F.M. Wiener, C.I. Malme, and C.M. Gogos. Sound propagation in urban areas. *Journal of acouctic society of America*, 37(4) :738–747, 1965.
- [Wri00] K. Wrightson. An introduction to acoustic ecology. *Soundscape*, 1(1) :10–13, 2000.
- [Xen63] I. Xenakis. Musiques formelles : nouveaux principes formels de composition musicale. *La revue musicale, édition spéciale*, 253-254 :1–233, 1963.

Résumé :

Les ressources mémoire et processeur allouées au rendu audio des applications vidéoludiques sont une contrainte importante dans le processus de réalisation d'ambiances sonores complexes. Des compromis sont nécessaires dans le choix des matériaux sonores et de leurs traitements pour parvenir à un résultat temps-réel crédible et immersif. Des méthodes alternatives dites *procédurales* tendent à être de plus en plus utilisées, elles consistent à modéliser le matériau sonore requis en fonction d'informations récupérées dans l'environnement virtuel. Les méthodes procédurales semblent appropriées à la sonorisation d'environnements virtuels complexes et évolutifs tels que des environnements urbains.

Dans ce travail de thèse, nous nous intéressons tout particulièrement à l'élaboration d'ambiances sonores urbaines interactives. Nous définissons la composition de ces ambiances au travers des analyses proposées par la théorie du Paysage Sonore et d'un état de l'art sur les applications vidéoludiques mettant en scène de vastes environnements urbains. Nous en retirons une segmentation du paysage urbain virtuel en sous-ensembles de plans sonores perceptifs, incluant un arrière-plan dont (première contribution) nous définissons les spécificités morphologiques et narratives. Nous considérons ainsi l'arrière-plan sonore urbain virtuel comme une *texture* et proposons (seconde contribution) de définir, spécifier et prototyper un outil de resynthèse granulaire dédié à sa sonorisation interactive.

Le prototype de synthétiseur, réalisé dans l'environnement de programmation graphique Pure Data. Sur la base de l'état de l'art, nous accompagnons cette proposition d'une méthodologie d'enregistrement des ambiances urbaines afin d'alimenter la resynthèse. Le prototype est validé en deux étapes : une intégration à la simulation de ville virtuelle Terra Dynamica un test perceptif d'écoutes comparatives.

Mots clefs :

synthèse granulaire, audio procédural, Pure Data, paysage sonore urbain, bruit de fond, jeu vidéo, design sonore, *field recording*, villes virtuelles

Abstract:

In video gaming and interactive media, the making of complex sound ambiances relies heavily on the allowed memory and computational resources. So a compromise solution is necessary regarding the choice of audio material and its treatment in order to reach immersive and credible real-time ambiances. Alternatively, the use of procedural audio techniques, i.e. the generation of audio content relatively to the data provided by the virtual scene, has increased in recent years. Procedural methodologies seem appropriate to sonify complex environments such as virtual cities.

In this thesis we specifically focus on the creation of interactive urban sound ambiances. Our analysis of these ambiances is based on the Soundscape theory and on a state of art on game oriented urban interactive applications. We infer that the virtual urban soundscape is made of several perceptive auditory grounds including a background. As a first contribution we define the morphological and narrative properties of such a background. We then consider the urban background sound as a *texture* and propose, as a second contribution, to pinpoint, specify and prototype a granular synthesis tool dedicated to interactive urban sound backgrounds.

The synthesizer prototype is created using the visual programming language Pure Data. On the basis of our state of the art, we include an urban ambiances recording methodology to feed the granular synthesis. Finally, two validation steps regarding the prototype are described: the integration to the virtual city simulation Terra Dynamica on the one side and a perceptive listening comparison test on the other.

Keywords:

granular synthesis, procedural audio, Pure Data, urban soundscape, background noise, video game, sound design, *field recording*, virtual cities